



武汉大学

本科生课程讲义及知识总结

课程名称：地理信息系统基础

开课学院：遥感信息工程学院

开课时间：2019-2020 年度第二学期

2020.武汉大学



目录

第一章·概述	4
➤ 地理信息系统的起因与发展	4
➤ GIS 概念内涵	4
➤ GIS 的 5 个基本特点以及分类:	7
第二章·地理信息系统的构成	9
➤ 地理信息系统的组成要素	9
➤ 地理信息系统的硬件构成	9
➤ 地理信息系统的软件构成	10
➤ 地理信息系统的主要功能	11
第三章·地理空间数据获取	11
➤ 空间数据	12
➤ 数据采集方式——野外数据采集	13
➤ 数据采集方式——地图数字化	14
➤ 数据采集方式——摄影测量数据采集	14
➤ 数据采集方式——遥感图像处理数据采集	15
➤ 数据采集方式——基于数字影像的地图数字化	16
➤ 数据采集方式——点云数据获取	16
➤ 数据采集方式——属性数据获取	16
➤ 数据采集方式——众源数据获取	17
➤ 空间数据质量	17
➤ 空间数据的元数据	19
第四章·空间数据表达	20
➤ 地理空间与地理现象	20
➤ 地理参考系统	20
➤ 地图投影	21
➤ 地理空间对象	24
➤ 空间对象关系	27
➤ 矢量数据表达	29
➤ 栅格数据表达	30
➤ 规则镶嵌结构	31
➤ 不规则镶嵌数据结构	33
第五章·地理空间数据处理	34
➤ 空间数据处理的基本算法	34
➤ 几何变换	37
➤ 空间数据转换	38
➤ 矢量与栅格的数据转换	40
➤ 拓扑关系的自动建立	43
➤ 矢量数据错误与编辑	45
➤ 图幅拼接与接边	47
第六章·地理空间数据处理	48
➤ 数据管理概述	48
➤ 空间数据组织	51

➤ 空间数据索引	52
➤ 空间数据库管理系统	54
➤ 空间数据查询	55
➤ 新型空间数据库系统	57
第七章·空间数据分析	59
➤ 矢量数据空间分析	59
➤ 栅格数据空间分析	63
➤ 三维数据空间分析	65
➤ 空间数据统计分析	70
第八章·空间数据可视化与地图制图	80
➤ 地学可视化	80
➤ 地图制图	81
➤ 地形数据可视化	86



第一章 · 概述

➤ 地理信息系统的起因与发展

1. **地理信息系统**：是一种特定而又十分重要的**空间信息系统**，它是以**采集、表达、处理、管理、分析和描述**整个或部分**地球表面**（包括大气层在内）与**空间和地理分布有关的数据的空间信息系统**。

2. GIS 发展的要素：

- 1) 以应用需求为驱动：地图、城市规划、城市管理、自然灾害风险评估等。
- 2) 以技术发展为导引：随着计算机技术的发展；摄影测量和遥感成像技术的发展等。

3. 国际上 GIS 的发展：

- 1) **20 世纪 50-60 年代（GIS 开拓期）**：①计算机水平不高，GIS 机助制图能力较强，能够实现地图的手扶跟踪数字化以及地图数据的拓扑编辑和分幅数据拼接功能；②基于格网系统，发展了许多基于栅格的操作方法；③专家以及政府推动，国际交流不多。

【1962 年加拿大提出利用计算机处理和分析大量的土地利用地图数据，于 1972 年建立了加拿大地理信息系统（CGIS），成为世界上第一个运行型地理信息系统】

- 2) **20 世纪 70 年代（GIS 巩固发展期）**：①GIS 朝实用方向发展；②大学和研究机构重视 GIS 软件设计和研究，GIS 专业人才不断增加；③技术发展没有新的突破；④专家影响减弱，政府影响增强。

- 3) **20 世纪 80 年代（GIS 普及和推广阶段）**：①GIS 开始注重空间决策支持分析；②GIS 应用领域迅速扩大。

- 4) **20 世纪 90 年代（步入网络 GIS 时代）**：①GIS 逐步进入网络 GIS 时代；②设计和开发在因特网上运行的 GIS 成为这个时期的主要趋势。

- 5) **21 世纪（GIS 逐步进入大众化应用时代）**：移动通讯的发展，将 GIS 发展为移动 GIS。

4. 我国 GIS 的发展：

- 1) 特点：起步稍晚，发展势头相当迅猛，已经成为世界上的 GIS 大国。
- 2) **20 世纪 70 年代（起步阶段）**：开始推广计算机在测量、制图和遥感领域中的应用。

【1977 年诞生了第一张由计算机输出的全要素地图】

- 3) **20 世纪 80 年代（试验阶段）**：全国性、区域性 GIS 建设，学术交流与人才培养取得了



大的发展。【1989 武汉测绘科技大学开设了 GIS 专业，成为全国第一个培养 GIS 本科人才的学校】

- 4) **20 世纪 90 年代（全面发展阶段）**：国家测绘局在全国范围内建立数字化测绘信息产业。
- 5) 二十一世纪向着集成化、产业化和社会化方向迈进。
 - A. 集成化：GIS 已成为一门综合性技术；
 - B. 产业化：GIS 产业化发展势头强劲；
 - C. 社会化：GIS 的社会化发展。

➤ GIS 概念内涵

1. **数据**：是未加工的原始资料，用以载荷信息的物理符号，数据本身并没有意义。
2. **信息**：用数字、文字、符号、语言、图形、图像等介质或载体，表示事件、事物、现象等的~~内容内容~~、数量或特征，向人们（或系统）提供关于现实世界新的事实和知识，作为生产、管理、经营、分析和决策的依据。

信息的特征：客观性、适用性、可传输性、共享性。

3. 数据和信息的区别和联系

- 1) 数据和信息密不可分；
 - 2) **信息来自于数据，数据是信息的载体**；
 - 3) **数据是未加工的原始资料**，数据是客观对象的表示，**信息则是数据内涵的意义**，是对数据的内容和解释；
 - 4) 只有理解了数据的含义，对数据做出了解释，才能提取数据中所包含的信息；
 - 5) **信息处理的实质是对数据进行处理，从而获得有用的信息。**
4. **地理数据**：是各种地理特征和现象之间关系的符号化表示。包括**空间位置特征、属性特征及时态特征**。

◇ **空间位置**：描述地理实体所在的**空间绝对位置**以及实体间空间关系的**相对位置**。

空间位置由坐标参照系统描述，空间关系由拓扑关系（邻接、关联、连通、包含、重叠等）描述。

◇ **属性特征**：又称为**非空间特征**，是地理实体的定性、定量指标，描述了地理信息的非空间组成成分。

◇ **时态特征**：指地理数据采集或地理现象发生的**时刻或时段**。时态特征正受到地理



信息系统学界的重视，成为研究热点

5. **地理信息**：指与所研究对象的空间地理分布有关的信息，是对表达地理特征和地理现象之间关系的地理数据的解释。
6. **地理信息的特性**：空间分布性、多维结构、时序特征、数据量大。
【地理信息的**空间位置、属性特征和时态特征**是地理信息系统技术发展的根本点，也是支持地理空间分析的三大基本要素】
7. **地理数据与地理信息的区别和联系**：
 - 1) 地理数据和地理信息是密不可分的；
 - 2) 地理信息来源于地理数据，地理数据是地理信息的载体，但并不就是地理信息；
 - 3) 地理信息处理的实质是对地理数据进行处理。
8. **空间认知的不同阶段**：数据(Data)——信息(Information)——知识(Knowledge)——智慧(Wisdom)，例如：32 度-温度 32 摄氏度-武汉 6 月平均气温 32 摄氏度-6 月开始进入夏天，病毒可能会自然死掉。
9. **地理实体**：指具有**地理空间参考位置**的地理实体特征要素，具有**相对固定**的空间位置和空间关系、**相对不变**的属性。特征要素分为离散特征要素、连续特征要素。
 - 1) **离散特征要素**：井、电力和通信线的杆塔、山峰的最高点、道路、河流、边界、市政管线、建筑物、土地利用和地表覆盖类型等；
 - 2) **连续特征要素**：温度、湿度、地形高程、植被指数、污染浓度等。
10. **地理现象**：指发生在地理空间中的地理事件特征要素，具有空间位置、空间关系和属性**随时间变化**的特性。
◇ 地理现象相对于地理实体的最典型区别：
地理现象是在一个**特定的时间段**存在的，具有一个发生、发展到消亡的过程；
地理实体具有**相对固定**的空间位置和空间关系、**相对不变**的属性
11. **地理对象**：是**地理实体和地理现象**在空间/时空信息系统中的**数字化表达形式**。地理对象包括**离散对象和连续对象**
12. **信息系统**：是对数据和信息进行**采集、存储、加工和再现**，并能**回答**用户一系列问题的**系统**。信息系统是**基于数据库的问答系统**。
13. **信息系统的组成**：信息系统由**计算机硬件、软件、数据和用户**四大要素组成的**问答系统**，智能化的系统还包括**知识**。【狭义的信息系统专指软件】



14. **信息的类型**：事务处理系统(Transaction Process System, TPS)、管理信息系统(Management Information System, MIS)、决策支持系统(Decision Support System, DSS)、人工智能和专家系统(Expert System, ES)。
15. **地理信息系统(GIS)的定义**：是一种特定而又十分重要的**空间信息系统**，它是以**采集、处理、管理、分析和描述**整个或部分**地球表面**（包括大气层在内）与**空间和地理分布**有关的**数据**的空间信息系统。
16. **GIS的发展**：GIS 已经从注重技术的**地理信息系统(GISystem, Geographic Information System)** 发展成理论与技术并重的**地理信息科学(GIScience, Geographic Information Science)**， 并逐步向侧重于服务的**地理信息服务(GIService, Geographic Information Service)**发展。
17. **地理信息科学(Geographical Informational Science, GIScience)**：是关于 GIS 的发展、使用和应用的理论，是**信息时代的地理学**，是关于**地理信息的本质特征与运动规律**的一门科学。
18. **地理信息服务**：遵循**服务体系的架构和标准**，采用**网络服务技术**，基于**地理信息互操作标准**和**规范**，在**网络环境**下提供地理信息系统数据、分析、可视化等功能的**服务**。
【软件就是服务】

➤ GIS 的 5 个基本特点以及分类：

1. **基本特点**：
 - 1) GIS 是以**计算机系统为支撑**的；
 - 2) GIS 的**操作对象是空间数据**；
 - 3) GIS 具有进行**空间分析、评价、可视化和模拟**的综合利用优势，具有**分析和辅助决策**作用；
 - 4) GIS **具有分布特性**；
 - 5) **GIS 的成功应用强调组织体系和人的因素的作用**。
2. **GIS 与相关系统的关系**：
 - 1) **GIS 与数字制图系统的区别与联系**：数字制图是地理信息系统的主要技术基础，主要区别在于空间分析，地理信息系统可以包含数字制图系统的所有功能，此外它还应具有丰富的空间分析功能。
 - 2) **GIS 与计算机辅助设计 (CAD) 的区别与联系**：二者都有坐标参考系统，都能描

述和处理图形数据及其空间关系，也都能处理非图形属性数据；CAD 处理的多为规则几何图形及其组合，图形功能极强，属性功能相对较弱；GIS 处理的多为地理空间的自然目标和人工目标，图形关系复杂；GIS 需要有较强的空间分析功能；CAD 一般仅在单幅图上操作。

- 3) GIS 与数据库管理系统的区别与联系：GIS 除需要功能强大的空间数据的管理功能之外，还需要具有图形数据的采集、空间数据的可视化和空间分析等功能。
- 4) GIS 与遥感图像处理系统的区别与联系：遥感图像处理系统是专门用于对遥感图像数据进行分析处理的软件，GIS 主要处理矢量数据。一些基于栅格的 GIS：除了能进行遥感图像处理，还具有空间叠置等 GIS 的分析功能，但是缺少空间关系描述；发展趋势：两者的进一步集成。

3. GIS 的分类：

- 1) 按范围大小：全球的、区域的和局部的；
- 2) 按维度：
 - A. 2-2.5 维：一般 GIS；
 - B. 真三维 GIS：布满整个三维空间建立的 GIS；
 - C. 2+1 维或 3 维 (2.5)：数字位置模型 (2 维) 和数字高程模型的结合；
 - D. 四维 GIS 或时态 GIS：加上时间。

4. GIS 与相关学科的关系：

①地理学为研究人类环境、功能、演化以及人地关系提供了认知理论和方法。②地图学为地理空间信息的表达提供载体与传输工具。③测量学、大地测量学、摄影测量与遥感等测绘学为获取这些地理信息提供了测绘手段。④宇航科学与技术为 GIS 向航空航天领域发展提供了新的理论和方法。⑤应用数学(包括运筹学、拓扑数学、概率论与数理统计等)为地理信息的计算提供数学基础。⑥系统工程为 GIS 的设计和系统集成提供方法论。⑦计算机图形学、数据库、数据结构等为数据的处理、存储管理和表示提供技术和方法。⑧软件工程、计算机语言为 GIS 软件设计提供方法和实现工具。⑨计算机网络、现代通信技术为 GIS 提供网络和通信的支撑技术。⑩人工智能、知识工程为 GIS 提供智能处理与分析的方法和技术。

5. 当代地理信息系统技术：组件 GIS、webGIS、多维动态 GIS、移动 GIS、实时 GIS、地理信息共享与互操作、地理空间信息公共服务【向公共开放的】等。<详见课件>



第二章 · 地理信息系统的构成

➤ 地理信息系统的组成要素

1. 从计算机的角度看：地理信息系统由**计算机硬件、软件(含空间分析)、数据和用户**四大要素组成。(硬件：软件：数据=1:2:7)
 - 1) **硬件**：GIS 的支撑，包括各类计算机处理机及其输入输出设备和网络设备；
 - 2) **软件 (包含空间分析)**：系统的功能驱动，是支持信息的采集、处理、存储管理、分析和可视化输出的计算机程序系统【**硬件和软件系统决定 GIS 的框架**】；
 - 3) **数据**：系统操作的对象，包括图形和非图形数据、定性和定量数据、影像数据及多媒体数据等【**数据处于核心地位**，数据在地理信息系统中具有重要地位。**数据是 GIS 的操作对象，是 GIS 的血液**，包括空间数据和属性数据。数据组织与管理的质量，直接影响 GIS 操作的有效性】；
 - 4) **用户**：**是地理信息系统所服务的对象**，是地理信息系统的主人。
2. **计算机的冯诺依曼体系**：计算机由计算器、控制器、存储器、输入设备和输出设备组成。
3. **要素寿命**：硬件寿命：一般较短；数据有效寿命最长，长的可以达到 70 年。
4. **地理信息系统的更新**：硬件更新、软件更新、数据更新、知识更新。

➤ 地理信息系统的硬件构成

1. **三种模式**：**单机模式、局域网模式、广域网模式**
 - 1) **单机模式**：**一台主机附带配置几种输入输出设备**，适用于 GIS 个别应用或小项目的应用。
 - 2) **局域网模式**：由计算机企业内部网、服务器集群、客户机群、磁盘存储系统（磁盘阵列）、输入设备、输出设备等支持的客户机/服务器（C/S）模式的 GIS 系统，**提供一个企业机构内的多用户共享操作服务**【内部用户通过企业内部网络实现数据资源、软硬件设备资源、计算资源的共享】
 - 3) **广域网模式（GIS 因特网系统模式，B/S 模式、B/S+C/S 混合模式）**：分布式计算模式：由因特网、服务器群、客户机群、磁盘存储系统（磁盘阵列）、输入设备、输出设备等支持的浏览器/服务器（B/S）模式的 GIS 系统，**提供因特网上许可用户的多用户操作**。【手机游戏属于广域网模式】



2. 单层结构和两层结构:

- 1) **单层结构**: 集中部署在一台独立的计算机设备上, 提供单用户使用系统的所有资源的一种方式, 以一台计算机为核心, 由输入设备、存储设备、输出设备共同组成的 GIS 硬件系统, 提供单用户操作;
- 2) **两层结构**: GIS 的资源和功能被适当地分配在服务器和客户机两端, 所有的客户端通过企业内部网, 共享网络资源, 进行信息共享和交换。

3. 计算机的核心部件: 中央处理器 (CPU) 和主存储器 (RAM)

主要性能指标: 主频、字长和内存容量, 网络配置和传输速度。

- 1) 中央处理器: 作用: 执行程序、控制所有硬件操作, 一台计算机可以安装多个。
- 2) 图形处理器 (GPU): 图形处理器(Graphics Processing Unit, GPU):一种专门在个人计算机、工作站和一些移动设备上处理绘图计算工作的微处理器。GPU 主要用于进行各种绘制计算机图形所需的运算。
- 3) 内存: 被 CPU 直接快速访问的存储区域。【内存和字长决定计算机的效率 (处理速度)】

4. 网络互连技术:

- 1) 网络互联形式: 同构型局域网的互连 (LAN-LAN); 异构型局域网的互连 (LAN-LAN); 局域网与广域网的互连 (LAN-WAN)。
- 2) 网络互连协议。
- 3) 网络互连设备。

5. 存储设备: 磁带、磁盘、可读光盘、可擦写光盘、硬盘、USB、云存储。

6. 输入设备:

- 1) 计算机通用输入设备: 键盘、光笔、鼠标;
- 2) GIS 专用输入设备: 数字化仪、扫描仪、摄影测量仪器、其他仪器、传感网设备、物联网设备。

7. 输出设备:

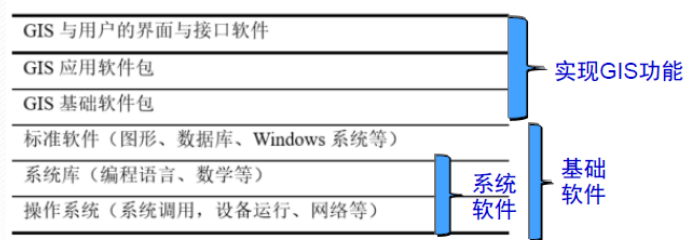
- 1) 通用输出设备: 打印机、绘图仪 (矢量绘图仪、栅格式绘图仪);
- 2) 图形显示终端: 用于图形的交互式输入、编辑、分析、处理、和输出。

➤ 地理信息系统的软件构成

1. GIS 产品的软件层结构:

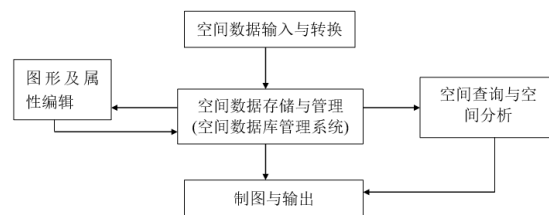
- 1) **系统软件**：与系统的硬件设备密切相关；（操作系统+系统库）
- 2) **基础软件**：系统软件连同标准软件共同组成保证 GIS 正常运行的基础软件；【此处的基础软件是维持计算机系统的正常运行】（系统软件+标准软件）
- 3) **实现 GIS 功能的软件**：**GIS 基础软件包、GIS 应用软件包、GIS 与用户的界面与接口软件**【GIS 应用软件包和 GIS 与用户的界面与接口软件是在基础软件包上面发展而来的，GIS 基础软件包几乎包括了 GIS 的所有功能】

表 2-3-1 GIS 产品的软件层结构



2. GIS 基础软件的五大子系统：

- 1) **空间数据输入与转换子系统；**
- 2) **图形及属性编辑子系统；**
- 3) **空间数据存储与管理子系统；**



【kml 是针对地理空间信息所开发的基于 xml 的标记语言】

- 4) **空间查询与空间分析子系统**【GIS 关心的重要内容】：
 - A. 位置问题：在特定的位置有什么或是什么；
 - B. 条件问题：符合某些条件的地理实体在哪里；
 - C. 变化趋势问题：某地理事件随时间变化的过程和趋势；
 - D. 模式问题：分析已发生或正在发生事件的相关原因；
 - E. 模拟问题：某个地区如果具备某种条件，会发生什么，过程如何？
- 5) **制图与输出子系统；**

➤ 地理信息系统的主要功能

地理空间数据获取、地理空间数据表达、地理空间数据处理、地理空间数据管理、地理空间数据分析、地理空间数据可视化与制图输出

第三章·地理空间数据获取

➤ 空间数据

1. **空间数据是 GIS 的血液**，（为什么说空间数据是 GIS 的血液）GIS 围绕空间数据的采集、表达、加工、存储、分析、可视化展开。

空间数据的定义：GIS 通过对地理世界进行抽象，并用数字表达。

2. **空间数据的类型（空间数据的基本内容）：**

- 1) **数字线划数据(Digital Line Graph, DLG)**：现有地形图上基础地理要素的**矢量数据集**，将空间地物抽象为**点、线、面等地理对象**，用**坐标**描述它们**位置和形状**，且保存空间地物间**空间关系**和相关的**属性信息**。【等高线是 DLG】

特点：①基于实体的数据；②拓扑关系复杂；③用抽象图形表达地物。

- 2) **数字栅格地形图(Digital Raster Graph, DRG)**：纸质地形图的数字化产品。每幅图经**扫描、纠正、图像处理及数据压缩**处理后，形成在内容、几何精度和色彩上**与地形图保持一致的栅格文件**

特点：①是**栅格图像**；②表面和线划图(DLG)一致，但实质不同；③通常作为某种信息系统的背景使用。

【栅格数据：无限放大之后是一个一个小格子；

矢量数据：无限放大之后不会有格状，由点线面组成。】

- 3) **影像数据(数字正射影像,Digital Orthograph Map, DOM)**：利用数字高程模型对扫描处理的数字化的航空像片/遥感影像，经过**像元纠正**，再按**影像镶嵌**，根据图幅范围剪裁生成的影像数据。

特点：①基于栅格表示的模型；②直观而详细记录地表自然现象；③数据源丰富；④生产效率高。

- 4) **数字高程模型(Digital Elevation Model, DEM)**：投影平面上规则格网点或三角网点平面坐标(x, y)及其高程(z)的数据集，用来表示地表物体的高程信息。

特点：

①DEM 的**水平间隔**可随地貌类型不同而改变

②根据**不同的高程精度**，可分为不同等级产品

③可以由**数字摄影测量**的方法全自动采集集得到

④可以采用**野外测量或地图扫描数字化**，经过内插处理得到

- 5) **属性数据(Attribute Data)**：描述空间地物的数量、质量、等级等特征的数据，是



GIS 的重要特征。【包含两层含义：有什么样的特性（定性描述）以及实体的详细描述信息（定量描述）】

3. 空间数据的基本特征：

- 1) **空间特征**：指空间地物的**位置、形状和大小等几何特征**，以及与相邻地物的**空间关系**。空间特征是**地理信息系统（空间信息系统）所独有的**。
- 2) **专题(属性)特征**：空间现象或空间目标的属性特征，是指除了时间和空间特征以外的空间现象的其它特征。
- 3) **时间特征**：空间数据总是在**特定的时间或时间段内**采集得到或计算得到的，时间特征是空间数据的重要特征之一。

4. 空间数据测量的尺度与精度：**比例尺可以反应空间数据测量的精度**

- 1) **空间尺度**：研究某一物体或现象时所采用的空间或时间单位，表示某一现象或过程在空间和时间上所涉及的范围和发生的频率。【尺度的参数：比例尺、精度、分辨率】

5. 空间数据的来源（数据是从哪里抽取的）：

- 1) 按照数据的内容：基础制图数据；②遥感图像数据；③数字高程数据；④自然资源数据；⑤调查统计数据；⑥法律文档数据；⑦多媒体数据；⑧已有系统数据；
- 2) 按照数据的来源：①非电子数据；②电子数据。

6. 数据获取方法：野外数据采集、地图数字化数据采集、摄影测量数据采集、遥感图像处理数据采集、点云数据采集、属性数据获取、众源地理数据获取

➤ 数据采集方式——野外数据采集

1. 早期模拟测量：

1) 平板测量：

A. 原理：**图解相似测量原理**；

B. 测量数据是非电子数据；

【将测量结果变成电子数据：手扶跟踪数字化、扫描数字化】

2) 经纬仪测量：

经纬仪：一种根据**测角原理**设计的测量水平角、竖直角和视距的测量仪器

经纬仪测量的缺点：测量强度大、测量精度低（相对于全站仪来说的）、测量效率低等。



2. 全站仪测量:

1) 全站仪: **电子经纬仪和激光测距仪**的集成, 可以同时测量空间目标的**距离和方位**数据, 并可进一步得到**大地坐标数据**。

2) 数据采集方式: ①测记法模式; ②电子平板模式

3. 全球导航卫星系统 (GNSS): 已经成为 GIS 的重要数据采集手段。

1) 定位原理: 接收机如果能够同时接收 4 颗以上卫星的信号, 根据三维空间后方交会原理, 由卫星的位置和接收机与卫星的距离, 计算出 GPS 接收机天线所在位置的 3 维地心坐标。

➤ 数据采集方式——地图数字化

1. 地图数字化: 将数据由**模拟格式**转化为**数字格式**的过程。

2. 地图数字化的方法: ①**手扶跟踪数字化**; ②**扫描数字化**。

1) 手扶跟踪数字化的缺点: 劳动强度大、精度相对较低、设备成本高;

2) 扫描数字化的两种方式:

A. 自动矢量化;

B. 交互式矢量化。

3. 地图数字化的特点: ①**简单、效率高**; ②**精度低**。

4. 地图数字化的关键: 建立设备 (图像) 坐标与地图坐标之间的**映射关系**, 也称为 (**图纸**) **定向**。

5. 地图数字化时应该注意: ①地物分层的正确性; ②地物符号应用的正确性; ③地物数字化方向的正确性。

6. 在扫描完之后需要进行**数字化编辑**: 数字化过程中, 由于存在设备误差、人为错误或原始地形图的错误, 导致数字化结果不可避免带有误差和错误。

7. 地图数字化可以获得包括: DLG、DRG 和 DEM

➤ 数据采集方式——摄影测量数据采集

1. 摄影测量的分类 (采集方式与特点):

1) 按**搭载平台**来分: 航空摄影测量、航天摄影测量和地面摄影测量

A. 地面摄影测量一般采用**倾斜摄影或交向摄影**;

B. 航空摄影测量一般采用**垂直摄影**。



- 2) 按处理技术手段来分：模拟摄影测量、解析摄影测量和数字摄影测量
 - A. 模拟摄影测量：利用模拟仪器；
 - B. 解析摄影测量：利用解析仪器；
 - C. 数字摄影测量：利用计算机。
2. 摄影测量有效的方式：立体摄影测量。

➤ 数据采集方式——遥感图像处理数据采集

1. 利用遥感的优点：
 - 1) 观测范围大；
 - 2) 能够提供大范围的瞬间静态图像、动态视频遥感影像；
 - 3) 能够进行大面积重复性观测；
 - 4) 光谱范围大，包含可见光、紫外线、红外线、微波等；
 - 5) 空间详细程度高，可以达到厘米级甚至是毫米级。
2. 遥感数据的分辨率：
 - 1) 空间分辨率：反映了数字遥感资料上对地物记录的详细程度；

【一般传感器的空间分辨率由其瞬时视场的大小决定，即由传感器内的感光探测器单元在某一特定的瞬间从一定空间范围内能接收到一定强度的能量而定，但一般使用其名义分辨率 = $\frac{\text{图像某行对应于地面的实际距离}}{\text{该行的像元数}}$ 】
 - 2) 光谱分辨率：指传感器所能记录的电磁波谱中某一特定的波长范围值，波长范围值越窄，光谱分辨率越高，传感器的波段数就越多；

☆光谱分辨率为 1nm 的比 10nm 的包含的谱线多
 - 3) 时间分辨率：指的是重复获取某一地区卫星图像的周期；

【一般而言：高时间分辨的影像其空间分辨率较低】
 - 4) 温度分辨率：是热红外遥感特有的指标，是指可以分辨率最小温度值
3. 遥感图像处理方式获取数据是 GIS 数据获取的重要方式之一。
4. 遥感数据处理流程：①观测数据输入；②再生、校正处理；③变化处理；④分类处理；⑤结果输出。
5. 遥感与摄影测量的关系：在含义上：遥感包括了摄影测量，在狭义上遥感和摄影测量的区别在于遥感的传感器的搭载平台比摄影测量更高。



➤ 数据采集方式——基于数字影像的地图数字化

1. **数字影像**：航天遥感影像、航空摄影影像及无人机遥感影像等

【特点：覆盖面积大、地物表达全、获取速度快、人员投入少等】

2. **基于数字影像的地图数字化的特点**：

- 1) 优点：工艺流程较简单、现实性强、设备硬软件投入较少、生产成本较低，特别适用于小区域、地形复杂地图生产与更新；
 - 2) 缺点：人机交互处理工作量较大、自动化程度较低、生产周期长、效率低
3. 从软件系统的发展来看，GIS 与遥感已经达到了“**无缝结合**”的阶段，将来的发展可能采取集成化程度更高的**整体结合方式**。

➤ 数据采集方式——点云数据获取

1. LiDAR：激光扫描与探测系统，简称为激光雷达。

2. **LiDAR 技术的优点**：

- 1) **主动性的工作方式**：不受日照和天气条件的限制，能够全天候实施对地观测；
- 2) **具有很强的穿透能力**：能部分地穿透树林遮挡，直接获取真实地面的高精度三维地形信息；
- 3) **作业安全**：能进行危险地区(沼泽、大型垃圾堆等)的测图工作；
- 4) **可以同其它技术手段集成使用**：提供新的融合数据源；
- 5) **作业周期快**：易于更新，时效性强，可很快提取测区 DEM；
- 6) 雷达没办法感应温度，主要是测量距离。

➤ 数据采集方式——属性数据获取

1. **属性数据包括**：

- 1) 目标类型的描述：地物类型定义；
- 2) 目标的具体说明与描述：宗地信息

2. **属性数据的获取方式**：资料调查与收集整理。

3. **属性数据的数据类型**：字符串、数字、图片、录像、声音、文本等

4. **属性数据的输入方式**：

- 1) 即时输入：对照图像直接输入；
- 2) 事后输入：一次性输入属性表，然后建立图形与属性之间的关系。



➤ 数据采集方式——众源数据获取

1. 众源地理数据：来源广泛，由非专业个人或单位生产的地理数据，大众用户成为数据和信息的提供者。
2. 众源地理数据处理与分析的关键技术：
 - 1) 众源地理数据的质量评价；
 - 2) 众源地理数据的信息提取与更新；
 - 3) 众源地理数据的分析与挖掘；
3. 众源数据获取的发展趋势：社会感知数据获取。
4. 众源数据的应用：广告推送、路段实时情况等。

➤ 空间数据质量

1. 数据质量：指数据适用于不同应用的能力。
2. 空间数据质量：地理数据正确反映现实世界空间对象的精度、一致性、完整性、现势性以及适应性的能力。
3. 空间数据质量的几个要素：
 - 1) 准确度(Accuracy)：测量值与真值之间的接近程度，用误差来衡量；
 - 2) 精度(Precision)：对现象描述的详细程度，用比例尺来衡量；
 - 3) 不确定性(Uncertainty)：当某种现象不能精确测定时，其值无法知道，其误差也就无法确定，这时就用不确定性代替误差；
【空间位置、属性、时域的不确定】
 - 4) 相容性(Compatibility)：两个来源的数据在同一应用中使用的难易程度；
 - 5) 一致性(Consistency)：对同一现象或同类现象的表达的一致程度；
【位置不一致和逻辑不一致】
 - 6) 完整性(Completeness)：具有同一准确度和精度的数据在特定空间范围内是否完整的程度；
 - 7) 可得性(Accessibility)：获取或使用数据的容易程度；
 - 8) 现势性(Timeliness)：数据反映客观现象目前状况的程度
4. 数据误差或不确定性的来源 【数据误差大小是一个累积的量】
 - ①按数据处理过程：

数据处理过程	误差来源
数据采集	野外测量误差：仪器误差、记录误差 遥感数据误差：辐射和几何纠正误差、信息提取误差 地图数据误差：原始数据误差、坐标转换、制图综合及印刷等误差
数据输入	数字化误差：仪器误差、操作误差 不同系统格式转换误差：栅格-矢量互换、三角网-等值线互换
数据存储	数值精度不够 空间精度不够：格网或图像太大、地图最小制图单元太大
数据处理	分类间隔不合理 多层数据叠加引起的误差传播：插值误差、多源数据综合分析误差 比例尺太大引起的误差
数据输出	输出设备不精确引起的误差 输出的媒介不稳定造成的误差
数据使用	对数据所包含信息的误解 对数据信息使用不当

②按误差特征：

误差特征	误差来源
明显误差	数据年代、地图比例尺、数据格式、数据的可接近性、数据代价
原始测量误差	位置误差、属性误差、数据输入输出误差、观测者偏差、获取数据时不同环境所引起的误差、自然变化
数据处理误差	计算机字长引起的误差、处理模型引起的误差、逻辑误差、地图叠置误差、拓扑关系所造成的误差、分类及处理方法引起的误差

③按误差类型：

地形图误差	地形图的位置误差、属性误差、时间误差、逻辑不一致性误差、不完整性误差
数据转换和处理误差	数字化误差（仪器误差、操作误差）、不同格式转换误差、不同GIS系统之间的转换误差
应用分析误差	数据层叠置时冗余多边形 数据应用时，应用模型引起的误差

5. 空间数据的误差类型：①几何误差；②属性误差；③时间误差；④逻辑误差

6. 几何误差及其描述：

1) 点误差：某点的测量值与真实位置之间的差异【采取点位误差分布椭圆表示】

2) 线误差：

A. 属性误差：产生于测量和对数据的后续处理的误差；

【道路、河流、市政或行政边界线】

B. 解释误差：在现实世界难以找到的线以及在确定线的界限时的产生的误差；

【数学投影定义的经纬线，按高程绘制的等高线】

【解译误差与属性误差直接相关，若没有属性误差，则可以认为那些类型界线是准确的，因而解译误差为零】

C. 直线误差与折线误差及其误差分布：

- i. 直线误差：一般以线的起点和终点处最大而中点误差最小；

直线误差的误差分布：**骨头型**；

- ii. 折线误差：一般在折线线段端点处较小，而在线段中点处较大；

折线误差的误差分布：**车链型**

- iii. 曲线误差的误差分布：**串肠型**

7. 属性误差：**命名、次序、间隔和比值**四种测度。

8. 属性误差的不确定性：**由属性的取值与其真值的相差程度决定**；

- 1) 连续属性数据误差：用量测误差来确定；
- 2) 类别属性数据误差：受多种因素影响。

➤ 空间数据的元数据

1. 元数据(Metadata)：关于数据的描述，是关于数据的数据。

2. 空间数据元数据：说明**地理空间数据**的数据。空间数据作为具有确切的地物位置信息、属性信息和时空信息的综合数据集，空间数据元数据就是对这些空间数据的详细描述或说明。

3. 空间数据元数据的标准：空间特征、属性特征、关系的描述

- 1) 标准是空间数据标准化的前提和保证
- 2) 建立标准才能有效利用空间数据

4. 元数据的作用：通过元数据可以检索、访问数据库，利用计算机系统资源，对数据进行加工和二次开发

- 1) 提供通过网络对数据进行查询、检索的方法或途径；
- 2) 帮助数据生产单位有效地管理和维护空间数据；
- 3) 提供有关数据生产单位数据存储、数据分类、数据内容、数据质量、数据交换网络及数据销售等方面的信息，便于用户查询、检索地理空间数据。

5. 空间元数据的应用：

- 1) 帮助用户获取数据；
- 2) 空间数据质量控制；
- 3) 在数据集成中有效地控制系统中的数据流；
- 4) 有利于数据存储和数据库功能的实现（可以避免数据的重复存储，可以高效查询

检索分布式存储数据)。

第四章 · 空间数据表达

➤ 地理空间与地理现象

1. **地理空间**：具有空间参考信息的地理实体或地理现象发生的时空位置集
2. **地理空间的分类**：
 - 1) 第一种分类：
 - A. 欧式空间：欧式距离公式；
 - B. 拓扑空间：拓扑关系的数学依据和基础。
 - 2) 第二种分类：
 - A. 绝对空间：空间坐标；
 - B. 相对空间：空间关系
3. **地理系统的特点**：
 - 1) 是一个**开放**的系统：和其他系统有关联、有交往，既有能量物质的交流，又有信息的交流。
 - 2) 是一个**复杂**的系统：由成千上万个子系统组成
4. **描述地理现象的四种几何类型**：
 - 1) 呈点状分布的地理现象；
 - 2) 呈线状分布的地理现象；
 - 3) 呈面状分布的地理现象；
 - 4) 呈体状分布的地理现象

➤ 地理参考系统

1. **大地水准面**：假定海水处于“完全”静止状态，把海水面延伸到大陆之下形成包围整个地球的连续表面【**不规则球面**】。
2. **地球椭球面**：一个大小和形状同大地水准面极为接近的**旋转椭球面**。
3. **地理坐标**：只能指经纬度【投影坐标不能当做地理坐标】。
4. **大地经纬度**：基准面是参考椭球面，基准线是法线；



- 1) 大地经度：指过参考椭球面上某一点的**大地子午面与起始子午面之间的二面角**；
- 2) 大地纬度：指过参考椭球面上某一点的**法线与赤道面的夹角**。
5. **天文经纬度**：以大地水准面和铅垂线为依据
 - 1) 天文经度 (λ)：起始子午面与过观测点的子午面所夹的二面角；
 - 2) 天文纬度 (φ)：过某点的**铅垂线与赤道平面之间的夹角**，铅垂线与法线不一定重合，它们之间的偏差称为**垂线偏差**。
6. **地心经纬度**：椭球面坐标系，基准面是参考椭球面
 - 1) 地心经度：等同于大地经度；
 - 2) 地心纬度：指参考椭球体面上的**任意一点和椭球体中心连线与赤道面之间的夹角**。
7. **空间直角坐标系【不属于地理坐标】**：右手直角坐标系 O-XYZ
 - 1) 原点：椭球体中心 O；
 - 2) X 轴：起始子午面与赤道面交线；
 - 3) Y 轴：赤道面上与 X 轴正交的方向；
 - 4) Z 轴：椭球体的旋转轴。
8. **我国常用坐标系**：1954 年北京坐标系、1980 年西安坐标系、2000 国家大地坐标系、WGS-84 坐标系。
9. **高程**：点到基准面的垂直距离
高程系统：1956 年黄海高程（1950-1956 年数据）、1985 国家高程基准（1952-1979 年）
10. **为什么日常不使用地理目标来描述位置**：用经纬度坐标不适合用于距离的测量。

➤ 地图投影

1. **地图投影的概念**：
 - 1) **数学上**：建立两个点集间一一对应的映射关系。
 - 2) **地图学上**：建立地球表面上的点（经纬度）与投影面上点（投影坐标）之间的一一对应关系。
2. **地图投影变形**：通过地图投影并按比例尺缩小绘制成的地图，存在**长度、面积和形状（角度）**的变化。
3. **长度变形**：用长度比来衡量。



长度比 (μ) : 投影面上的一微小线段与地球椭球面上相应的微小线段之比。

长度变性值: $V\mu = \mu - 1$

☆正值: 投影后长度增加; 负值: 投影后长度变短; 0: 投影后无变形

☆中央经线上没有长度变形, 随着经差的增加经线的长度变形增大

4. 面积变形: 用面积比和面积相对变形值来表示

面积比 (P): 投影平面上微小面积与地球椭球面上相应的微小面积之比。

☆纬度越高, 面积比越大, 相对变形值就大

相对变形值 (Vp): 以面积的相对变形衡量面积变形的大小 $Vp = P - 1$

☆ $Vp > 0$: 表示投影后面积增大; $Vp < 0$: 表示投影后面积缩小;

$Vp = 0$: 表示投影后面积不变

5. 角度变形: 投影面上任意两方向线所夹的角度与地球椭球面上相应两方向线夹角之差

☆角度变形可以做到全图角度不变形, 但同时长度和面积会变形

☆只有中央经线和各纬线相交成直角, 其余的经线和纬线均不呈直角相交

6. 地图投影类型: 方位投影、圆锥投影、圆柱投影

◇ 正轴投影、斜轴投影、横轴投影; 相切投影、相割投影

1) 方位投影: 用平面与地球模型相切或相割而将球面或椭球面上的点转换到平面上的投影叫方位投影。

✓ 投影面的形状是平面, 投影完之后纬线变成了同心圆。

2) 圆锥投影: 将球面或椭球面上的点置换到锥面上的投影。

✓ 投影完纬线变成了同心弧线

3) 圆柱投影: 将球面或椭球面上的点转换到柱面上的投影叫圆柱投影。

✓ 圆柱投影是圆锥投影的特例, 圆锥投影的性质完全适用于圆柱投影, 反之不成立

7. 我国常用的地图投影:

1) 大中比例尺 (1:50 万以上): 高斯-克吕格投影 (横轴等角切椭圆柱投影)

2) 小比例尺 (1:100 万以下): Lambert 投影/兰勃特投影 (正轴等角割圆锥投影)

3) 墨卡托投影 (正轴等角圆柱投影)

4) UTM 投影 (通用横轴墨卡托投影)

8. 高斯-克吕格投影:



1) 投影条件 (如何进行投影):

- A. 中央经线和地球赤道投影成为直线且为投影的对称轴;
- B. 等角投影;
- C. 中央经线上没有长度变形;

2) 投影特点:

- A. 中央经线上无变形;
- B. 同一条纬线上, 离中央经线越远, 变形越大
- C. 同一条经线上, 纬度越低, 变形越大
- D. 等变形线为平行于中央经线的直线
- E. 为了保证精度, 常采用分带投影:

☆小比例尺用 6° 分带 【1:2.5 万至 1:50 万的地形图均采用 6° 分带方案】;

☆大比例尺用 3° 分带

9. 兰勃特投影(正轴等角割圆锥)的特点:

- 1) 角度没有变形;
- 2) 两条标准纬线上没有任何变形;
- 3) 等变形和纬线一致, 同一条纬线上的变形处处相等;
- 4) 在同一经线上, 两标准纬线外侧为正变形 (长度比大于 1), 两标准纬线之间为负变形 (长度比小于 1);
- 5) 同一纬线上等经差的线段长度相等, 两条纬线间的经线长度处处相等。

10. 墨卡托投影的特点:

☆常用于航海和航空, 在地图上保持方向和角度的正确。

- 1) 没有角度变形
- 2) 由每一点向各方向的长度比相等
- 3) 经纬线都是平行直线, 且相交成直角, 经线间隔相等, 纬线间隔从标准纬线向两极逐渐增大

11. UTM 投影 (通用横轴墨卡托投影) 的特点:

- 1) 中央子午线长度变形比为 0.9996 【其他三种的中央子午线没有长度变形】
- 2) 两条割线上没有变形, 离开这两条割线越远则变形越大
两条割线以内: 长度变为负值 两条割线以外: 长度变为正值

12. 高斯投影与 UTM 投影异同：

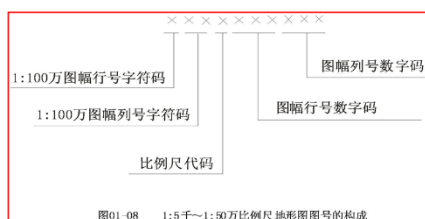
- 1) 比例因子：高斯-克吕格投影中央经线上的比例系数为 1，UTM 投影为 0.9996
- 2) 分带方式：两者的分带起点不同
 - A. 高斯-克吕格投影自 0 度子午线起每隔经差 6 度自西向东分带，第 1 带的中央经度为 3°
 - B. UTM 投影自西经 180° 起每隔经差 6 度自西向东分带，第 1 带的中央经度为 -177° ，因此高斯-克吕格投影的第 1 带是 UTM 的第 31 带
- 3) 两投影的东伪偏移都是 500 公里。

13. 地形图的分幅和编号：

- 1) 分幅方法：
 - A. 按经纬线划分的梯形分幅法→用于国家基本比例尺系列的地形图
 - B. 按平面直角坐标格网划分的矩形分幅法→大比例尺地形图
- 2) 我国采用梯形分幅法，统一按经纬度划分。
- 3) 分幅编号：以 1:100 万地形图为国家基本比例尺地形图分幅和编号

☆采用国际统一的行列式编号方法：

 - A. 横行：从赤道起算，每纬差 4° 为一行，至南、北纬 88° 各分为 22 行，依次用大写拉丁字母 A、B、C、...V 表示行号；
 - B. 纵列：从 180° 经线起算，自西向东每经差 6° 为一列，全球分为 60 列，依次用阿拉伯数字 1、2、3、...60 表示其相应列号



比例尺	1:50万	1:25万	1:10万	1:5万	1:2.5万	1:1万	1:5千
代 码	B	C	D	E	F	G	H

➤ 地理空间对象

1. 空间数据模型：按照一定的规则（模型）来建立描述客观实体本身以及实体之间联系的模型
2. 空间数据模型分类：
 - 1) 基于对象（object-based）的空间数据模型

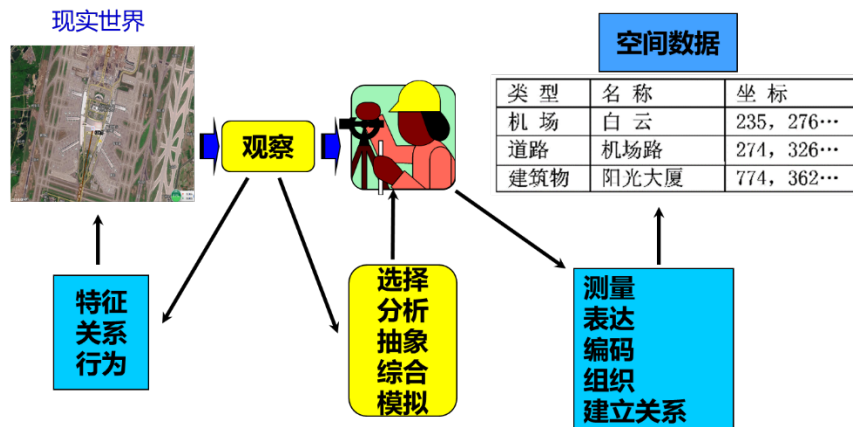
2) 基于域 (field-based) 的空间数据模型

☆在计算机中场模型是用栅格 (Raster) 数据结构来实现

3. 地理空间对象: 地理实体和地理现象在空间(时空)信息系统中的数字化表达形式。

- 1) 地理实体和地理现象是在现实世界中客观存在的;
- 2) 地理空间对象是地理实体和地理现象的数字化表达。

4. 地理空间对象的抽象过程:



首先, 通过对现实地理世界的观察, 分析其**特征、关系和行为**, 进一步通过**选择、分析、抽象、综合和模拟**等, 实现现实地理世界的数字化表达, 得到地理空间对象, 通过对地理空间对象进行**测量、表达、编码、组织、建立关系**等, 得到地理空间对象的属性特征。

5. 地理空间对象的描述: 识别码、位置、空间特征、行为和功能

- 1) 识别码: 用于区别同类而又不同的实体, 通常用对象标识码 (OID) 来表示, 如同为一级公路类型的不同的公路;
- 2) 位置: 可用坐标描述, 也可用其它形式 (如邮政编码);
- 3) 空间特征: 也是位置信息的一种, 如维数、类型及实体的组合。地理空间对象的维数可以分零维、一维、二维、三维、甚至四维 (三维 + 时间)。
- 4) 行为和功能: 是指在数据采集过程中不仅要重视静态描述, 还要收集那些动态的变化, 如岛屿的侵蚀、水体污染的扩散、建筑的变形等。

6. 地理空间对象的分类:

- 1) **零维空间对象**: 点状空间实体, 由坐标 (x, y, z) 定义其位置
 - A. 实体点: 用来代表一个实体
 - B. 标识点 (注记点): 用于定位地图或插图的文本信息 (注记)

- C. 面标识点：用于负载多边形的属性，存在于多边形内
 - D. 节点：表示线（或弧）的终点和起点
 - E. 顶点：表示线段和弧段的内部点
- 2) **一维空间对象**：线状空间实体，由一系列有序坐标串表示
- A. 线段：两点之间的直线段；
 - B. 弦列：相互连接无分支的线段；
 - C. 弧：曲线轨迹，可以用数学函数定义；
 - D. 拓扑连线：两个节点之间的拓扑连接，由结点的顺序确定其方向；
 - E. 链：非相交的线段或弧的无分支而有向的序列
 - a) 全链：可以显式定位左右多边形和始终端节点的链；
 - b) 面链：可以显式定位左右多边形、但不能定位始终端节点的链；
 - c) 网链：可以显式定位始终端节点、但不能定位左右多边形的链
 - d) 环：是一个由不相交的链或弦列和（或）弧的闭合系列，
 - F. 它只表示一个封闭的边界，但不表示封闭内的面积。
 - a) G-环：边界系列无方向
 - b) GT-环：边界系列有方向，由多条链组成
- 3) **二维空间对象**：面状空间实体，在数据库中由一封闭曲线加内点来表示。
- A. 内面：不包括其边界的面；
 - B. G-多边形：由一个内面、一个外 G-环和零个或多个不相交的内 G-环组成的面；
 - C. GT-多边形：由 GT-环为边界定义的多边形，也可以由多个链直接定义；
 - D. 广义多边形：定义为 GT-多边形覆盖的面的周边以外的面；
 - E. 虚多边形：为二维拓扑面的一部分，以其它的 GT-多边形为界，但其它方面有与广义多边形相同的特性；
 - F. 像元：是一个二维的图素，是不可再分的图像元素；
 - G. 格网单元：一个网单元的二维对象。
- 4) **三维空间对象**：体状空间实体，是三维空间中的现象与物体
- A. 体元：方形实体，三维实体中不可再分割的元素
 - B. 标识体元：用于标识一个三维空间，类似于多边形中的标识点

- C. 三维组合空间目标：由二维空间目标组合（房屋是由多个面组合而成）、也可以由三维体元构成
 - D. 体空间：三维空间实体，它可以划分为体元表示，即为体元的聚集目标；它也可以是由不规则表面表示的体空间。
- 5) **聚合空间对象**：由不同的空间单元组合而成
- A. 格网：某种面的规则(矩形、三角形)或接近规则的镶嵌拼接
 - B. 图像：像元集合
 - C. 层：表示某一种专题的空间对象的集合，如道路层、水系层、植被层
 - D. 栅格：同一类型网格或图像的一系列覆盖层
 - E. 图形：相互关联的零维、一维或二维对象的集合。
 - F. 平面图形：组成图形的节点、拓扑连线或链对象在平面上的图形，在同一个平面上；
 - G. 二维拓扑面：指一个平面图形及相关联的二维对象；
 - H. 网络：由节点、拓扑连线组成，可以不在一个平面上。通过路径连接多个地理现象之间的相互关系。

➤ 空间对象关系

1. **空间对象关系**：地理实体之间存在的与空间特性有关的关系。
三种基本类型：**拓扑关系、方位关系和度量关系**，**拓扑关系是最重要的空间关系**
2. **拓扑关系**：指空间对象在拓扑变换（旋转、平移、缩放等）下保持不变的空间关系，即拓扑不变量。
3. **拓扑关系的类型**：
 - 1) **重合关系**：基本几何体与比较几何体完全重合；
☆只有相同维数的空间目标才存在重合关系
 - 2) **包含关系**：指基本几何体包含比较几何体，是基本几何体的一部分；
☆只有高维目标包含低维目标，或者同维目标包含同维目标，不能包含比自身维数高的几何体
 - 3) **位于内部关系**：与包含关系相对，即基本几何体处于比较几何体之内；
 - 4) **相交关系**：**线与线之间、线与面之间，或者两个多边形间存在交叉**；
 - 5) **相离关系**：指基本几何体与比较几何体无共享点；



6) **重叠关系**：指基本几何体覆盖比较几何体；

☆只有相同维数的空间目标才存在重叠关系

7) **相切关系(邻接关系)**：指两个几何体具有公共边或点。

4. **拓扑关系的描述**：4 交模型 (Egenhofer & Franzosa,1991)、9 交模型 (Egenhofer & Herring,1991)、RCC 模型 (Region Connection Calculus, 区域连接演算) (Randell et al.,1992)、空间代数模型 (Li, 2001)、V9I 模型 (陈军,2002) 等。
5. **4 元组模型**：又名 4 交模型，是一种基于点集拓扑学的二值拓扑关系模型。将每个空间实体 A 表示为由边界点集 ∂A 和内部点集 A^o 构成的集合，两个空间实体间的拓扑关系通过物体 A 的边界 (∂A) 和内部点 (A^o) 与物体 B 的边界 (∂B) 和内部点 (B^o) 间的交集来描述：

$$T(A, B) = \begin{bmatrix} \partial A \cap \partial B & \partial A \cap B^o \\ A^o \cap \partial B & A^o \cap B^o \end{bmatrix}$$

6. **9 元组模型**：针对 4 交模型缺陷，9 交模型引入空间实体的“外部”，构成 9 元组模型。

$$R(A, B) = \begin{bmatrix} \partial A \cap \partial B & \partial A \cap B^o & \partial A \cap B^- \\ A^o \cap \partial B & A^o \cap B^o & A^o \cap B^- \\ A^- \cap \partial B & A^- \cap B^o & A^- \cap B^- \end{bmatrix}$$

7. **方向关系**：指源目标相对于参考目标的顺序关系(左右前后、东南西北等)

两种描述方式：定性方向描述、定量方向描述。

代表性模型：锥形模型、方向关系矩阵模型等。

- 1) 定量描述模型：主要是以角度为参数，实现方向关系定义和描述；
- 2) 定性方向描述：利用东南西北等相对参考，广泛应用于人们日常生活中相对定位
- 3) 锥形模型：将空间目标及其周围的区域划分为带有方向性的互斥的若干个锥形区域（一般为 4 个或者 8 个），利用源目标与方向区域取交结果来定义方向关系。
- 4) 方向关系矩阵模型：利用参考目标的 MBR(Minimum Boundary Rectangle, 最小外接矩形)把平面空间划分为九个方向片，将源目标与这九个方向片分别求交，得到方向关系矩阵。

8. **距离关系**：

- 1) **欧式距离**： $d(A, B) = ||A - B|| = [\sum_{i=1}^n (a_i - b_i)]^{1/2}$



- 2) 曼哈顿距离 (Manhattan): 又名绝对值距离、街坊距离: $d(A, B) = \sum_{i=1}^n |a_i - b_i|$
- 3) 切比雪夫距离: $d(A, B) = \max_i |a_i - b_i|$
- 4) 闵可夫斯基(Minkowski)距离: $d(A, B) = [\sum_{i=1}^n |a_i - b_i|^m]^{1/m}$
- 5) 马氏距离
- 6) 汉明距离
- 7) 大地测量距离
- 8) 定性距离关系描述: 邻近关系: 远、中、近等

➤ 矢量数据表达

1. 矢量数据模型: 采用 **矢量数据结构** 来描述, 是基于实体的数据结构, 通过记录空间对象的坐标(x, y, z)及空间关系来表达空间对象的位置。
2. 矢量数据结构表达的内容: 矢量数据自身的存储和管理、几何数据和属性数据的联系、空间对象的空间关系 (拓扑关系)。
3. 空间关系表达: 空间关系表达 = 信息结构 + 操作。
4. 矢量数据结构的基本元素:
 - 1) 点实体: 存储点实体的 x, y 坐标, 以及一些与实体有关的数据;
 - 2) 线实体: 唯一标识码、线标识码、起始点、终止点、坐标点对系列、非几何信息等;
 - 3) 多边形数据结构: 既要表示位置和属性, 还要表达区域的拓扑特征 (形状、邻域或层次结构)。
5. 无拓扑关系的矢量数据模型: 也称面条数据模型(Spaghetti), 仅记录空间对象的位置坐标和属性信息、不记录拓扑关系。
 - 1) 存储方式:
 - A. 独立存储 (多边形环路法): 空间对象位置直接跟随空间对象
 - B. 点位字典法: 点坐标独立存储, 线、面由点号组成
 - 2) 特点:
 - A. 无拓扑关系, 主要用于显示、输出及一般查询
 - B. 公共边重复存储, 存在数据冗余, 难以保证数据独立性和一致性
 - C. 多边形分解和合并不易进行, 邻域处理较复杂

D. 处理嵌套多边形比较麻烦

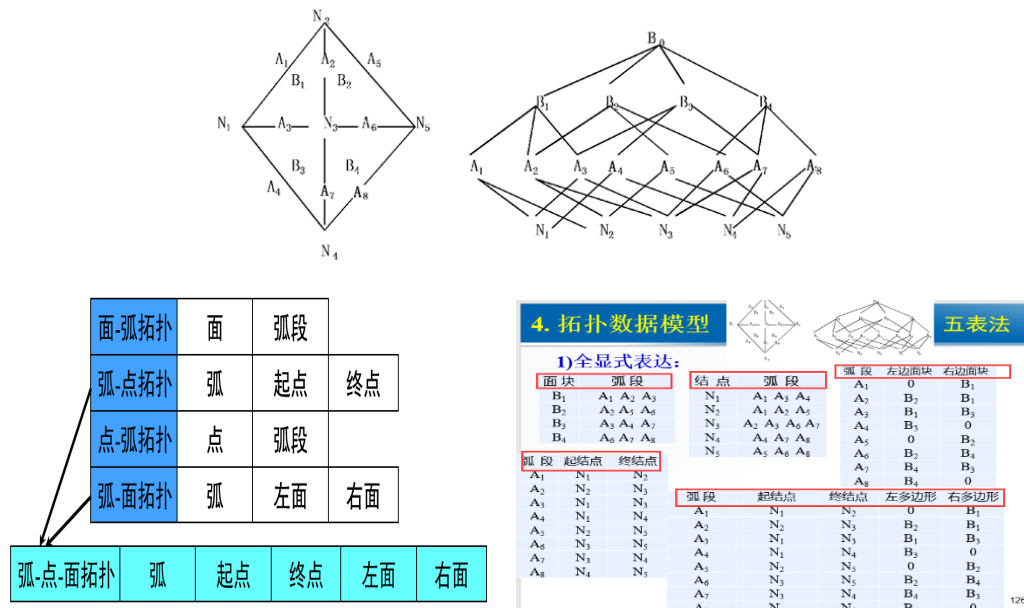
E. 适用于制图和一般查询，不适合复杂的空间分析

6. **拓扑数据模型**：不仅表达空间对象的几何位置和属性，还表示空间关系

表达方式：全显式表达、部分显式表达、拓扑关系与数据共享

7. **全显式表达**：结点、弧段、面块之间的所有关联拓扑关系都用关系表显式地表达出来

→ ☆五表法【记住五张表的表头，给图能填表】



8. **部分显式表达**：仅表示结点、弧段、面块之间部分拓扑关系，是全显式表达的子集。

9. **属性特征类型**：类别特征：是什么+说明信息：解决两个同类目标的不同特征问题。

10. **属性特征表达**：类别特征：类型编码+说明信息：属性数据结构和表格

➤ 栅格数据表达

1. **栅格数据结构**：是基于位置的数据结构，采用像元阵列来描述空间对象，每个像元的行列号确定位置，用像元值表示空间对象的类型、等级等特征。

2. **栅格数据**：是二维表面上地理数据的离散量化值（是地理数据的近似值）

☆每个栅格单元只能存在一个值，像元值是近似值，具有相邻的四个像元组成的正方形可以是最小的像元。

3. **层**：当同一像元要表示多种地理属性时，则需要多个笛卡儿平面来描述，每个笛卡儿平面表示一种地理属性或同一属性的不同特征，这个平面就是“层”

☆栅格数据可以用平面笛卡儿空间来描述，但一个平面笛卡儿空间只能具有一个属性数据，每一层对应一个栅格文件。



4. 栅格数据的组织方法:

- 1) 方法一: 以像元为记录的序列, 不同层上同一像元位置上的不同属性值表示为一个列数组。

☆节省存储空间 (像元坐标空间)

- 2) 方法二: 以层为基础, 每一层记录后再记录下一层。
- 3) 方法三: 以层为基础, 但每一层则以多边形为序记录多边形的属性值和充满多边形的各象元的坐标。

5. 网格系统: 是栅格系统的一种特殊形式, 也是基于位置的数据模型, 适用于多元、多层次的数据管理系统。

6. 网格系统与栅格系统的区别:

- 1) 网格单元可以独立地存取, 栅格则扫描整行;
- 2) 栅格数据库一般呈矩阵状, 而网格数据库可呈不规则状

➤ 规则镶嵌结构

1. 镶嵌结构: 表达空间域或者称场对象的空间数据结构。
2. 规则镶嵌数据结构: 用规则的小面块集合来逼近自然界不规则的地理空间。
3. 格网单元: 具有简单的形状和平移不变性。
4. 格网系统: 可以由正方形镶嵌数据结构构成。将矩形格网叠置在研究区域, 把连续的地理空间离散为互不覆盖的面块单元。

5. 四叉树分割的基本思想:

- 1) 首先把一幅图像 (栅格地图) 等分成 4 部分;
- 2) 如果某子区所有格网都含有相同值, 就不再往下分割;
- 3) 否则, 再分割成四个子区域;
- 4) 递归地分割, 直至每个子块都只含有相同值为止

6. 常规四叉树: 除了记录叶结点, 还要记录中间结点, 结点之间的联系靠指针表达, 也叫指针四叉树。

☆常规四叉树靠指针来维护父子之间的相互关系, 增加了存储量和操作的复杂性

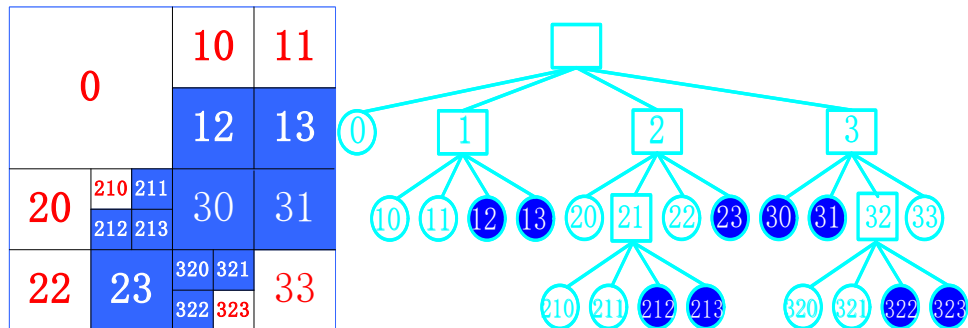
☆每一个节点包括: 父节点指针、4 个子节点指针、本节点灰度或属性值

7. 线性四叉树: 只存储最后叶结点的信息, 即结点的位置、大小和格网值, 叶结点的编号遵照一定的规则, 这种编号称为地址码, 它隐含了叶结点的位置信息。

叶结点位置：基于四进制或十进制的 Morton 码表示

8. 基于四进制的 Morton 码及四叉树的建立：

1) 自上而下分裂方式：自上而下，在建立四叉树的过程中逐步产生 Morton 码；



2) 自下而上的合并方式：先计算每个格网的 Morton 码，然后按一定的扫描方式采用自下而上的合并方法建立四叉树【效率比第一种方式高】

A. 第一步：根据行列号，并将十进制的行列号转换成二进制树

B. 第二步：计算对应的 Morton 码 MQ，并将元素按码的升序排列成线性表

$$M_Q = 2 \times I_b + J_b$$

M _Q 行号		列号	J _b							
			JJ	0	1	2	3	4	5	6
II	I _b		0	1	10	11	100	101	110	111
0	0		000	001	010	011	100	101	110	111
1	1		002	003	012	013	102	103	112	113
2	10		020	021	030	031	120	121	130	131
3	11		022	023	032	033	122	123	132	133
4	100		200	201	210	211	300	301	310	311
5	101		202	203	212	213	302	303	312	313
6	110		220	221	230	231	320	321	330	331
7	111		222	223	232	233	322	323	332	333

C. 第三步：依次检查每四个相邻 MQ 码对应的格网值，如果相同则合并为一个块，否则将这四个格网记录下来，内容包括 MQ 码、深度和格网值。

0			10		11	
			12		13	
20	210	211	30			31
	212	213				
22	23		320	321	33	
			322	323		

D. 第四步：第一轮检测完以后依次检查每相邻四个大块的格网值，若其中有一个值不同或某子块已记录，则不作合并而记录，循环下去直到没有能够合并的子块为止。

9. 基于十进制的 Morton 码及四叉树的建立：采用十进制的 Morton 码作为线性四叉树的地址码，并且使用自下而上的合并方法建立四叉树。

1) 将十进制的行号和列号转换为伪码(I_f, J_f)



方法：将十进制的行列号转换成二进制，从最低位开始依次按位乘 4^n ，

$$\text{例如 } II = (6)_{10} \xrightarrow{\text{转 2 进制}} 110 \xrightarrow{\text{按位乘 } 4^n} I_f = 0 \times 4^0 + 1 \times 4^1 + 1 \times 4^2$$

2) 计算对应的 Morton 码 M_D : $M_D = 2 \times I_f + J_f$

M _D 码 行号 \ 列号		JJ	0	1	2	3	4	5	6	7
		J _f	0	1	4	5	16	17	20	21
II	I _f									
0	0		0	1	4	5	16	17	20	21
1	1		2	3	6	7	18	19	22	23
2	4		8	9	12	13	24	25	28	29
3	5		10	11	14	15	26	27	30	31
4	16		32	33	36	37	48	49	52	53
5	17		34	35	38	39	50	51	54	55
6	20		40	41	44	45	56	57	60	61
7	21		42	43	46	47	58	59	62	63

3) 进行合并操作：依次检查每四个相邻 Morton 码对应的属性值，如果相同则合并为一个块；

4) ☆如果在相邻区域内有**顺序连续**的三个相邻的 Morton 码对应的属性值一致，也可以将其合并【**但是不建议，合并了之后就不再是 4 叉树了**】，以从左上角开始按**顺序（例如 0-1-2-3）**遇到的第一个 Morton 码代表这三个 Morton 码。【注意与四进制区分：四进制必须四个块都相同】

10. **二维行程编码**：在线性四叉树中，仍然存在前后叶结点的值相同的情况，因而可以采取进一步的压缩表达，即将格网值相同的前后结点合并成一个值，形成新的线性列表

11. **二维行程编码的特点**：

- 1) 前后两个地址码之差表达了该行程码的格网数，它可以直接表示该子块的大小；
- 2) 压缩效率更高、节省存储空间，方便以后的插入、删除和修改等操作

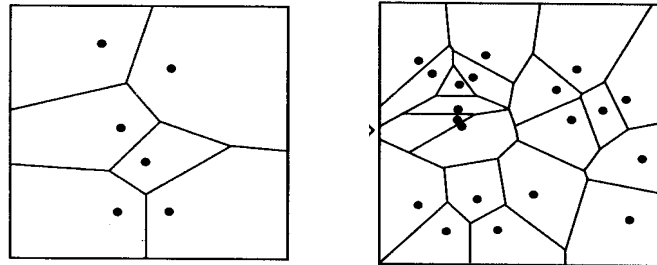
12. **矢量栅格合一的一体化数据结构**：具有矢量实体的概念，同时又具有栅格覆盖的思想，具有栅格矢量两种数据结构的特点。

➤ 不规则镶嵌数据结构

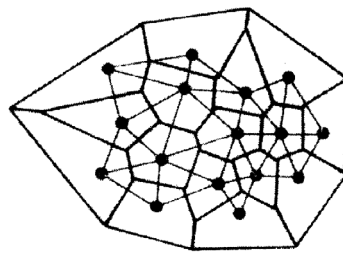
1. **不规则镶嵌数据结构**：是指用来进行镶嵌的小面块具有不规则的形状或边界。
2. **不规则镶嵌数据结构的种类**：Voronoi 图(泰森多边形)、狄洛尼(Delaunay)三角网、不规则三角网 TIN——数字高程模型



3. **Voronoi 图(泰森多边形):** 设有平面点集 $S(P_1, P_2, \dots, P_n)$, 其对应的 Voronoi 多边形 $V(P_1), V(P_2), \dots, V(P_n)$, 其中 $V(P_i)$ 是由距 P_i 最近的所有点组成的, 属于 $V(P_i)$ 的每一个点 x 到 P_i 比到 S 的任何其它点 P_j ($i \neq j$) 都近, 即 $V(P_i)$ 的内部是到 P_i 点比到 S 的其余点更近的全部点的集合。



4. **Delaunay 三角网:** 可以由 Voronoi 图中构造, 也可以根据离散点直接构造。是 Voronoi 图的对偶, 将 Voronoi 图中各多边形单元的内点连接后得到一个布满整个区域而又互不重叠的三角网结构。



第五章 · 地理空间数据处理

➤ 空间数据处理的基本算法

1. 点状数据处理基本算法:

点到点距离:

$$\text{两点间的直线距离: } d = |AB| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

$$\text{两点间的球面距离: } d = R \times \arccos(\sin x_1 \sin x_2 + \cos x_1 \cos x_2 \cos(y_1 - y_2))$$

点到线距离:

设有一直线段 L , 两端点的坐标为 (x_A, y_A) 和 (x_B, y_B) , 另一给定点 P 的坐标 (x_P, y_P)

点P到直线L的线距离D为：

$$D = |y_B x_P - y_A x_P + x_A y_P - x_B y_P + y_A x_B - x_A y_B| \sqrt{(y_B - y_A)^2 + (x_A - x_B)^2}$$

点到一条线目标（包含多条折线段）的距离：

1) 方法1：求点到线目标各条折线段内的距离(不是延长线上的距离)：距离最小的值即为点到该线目标的距离。

2) 方法2：计算点到一个线状目标顶点的距离，然后取其最小值

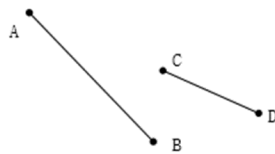
2. 线状数据处理基本算法：

线目标的几何计算：主要为多个线段组成的线目标的长度计算。

矢量线的长度为组成矢量线的线段长度之和 $L = \sum_{i=1}^n d_i$

拓扑关系判断：

两条线段AB和CD，端点坐标分别为 (x_A, y_A) , (x_B, y_B) , (x_C, y_C) , (x_D, y_D)



$$\left. \begin{aligned} t &= \frac{(x_C - x_A)(y_C - y_D) - (x_C - x_D)(y_C - y_A)}{(x_B - x_A)(y_C - y_D) - (x_C - x_D)(y_B - y_A)} \\ s &= \frac{(x_B - x_A)(y_C - y_A) - (x_C - x_A)(y_B - y_A)}{(x_B - x_A)(y_C - y_D) - (x_C - x_D)(y_B - y_A)} \end{aligned} \right\}$$

t为线段AB的参数，s为线段DC的参数，如果t和s均在0和1之间，则存在交点

曲线光滑：线性迭代光滑法（抹角法）、二次多项式平均加权插值法、五点求导分段三次多项式插值法、张力样条函数。

曲线化简：与曲线光滑相反，曲线化简是要删除或者说压缩一部分曲线上的采样点。

平行线处理：直线段、折线、曲线。

直角平差处理：由于测量误差，使得某些本来垂直的直线段互相不垂直。

3. 多边形基本操作算法：

几何计算：

多边形周长：多边形目标周长是所有线段两点之间距离的总和。

多边形面积：设多边形S是由顶点P₁, ..., P_n组成的多边形，则S的面积为：

$$S = \frac{1}{2} \left(\sum_{i=1}^{n-2} (x_i y_{i+1} - x_{i+1} y_i) + (x_n y_1 - x_1 y_n) \right)$$

多边形的重心：重心是分别求多边形顶点x坐标和y坐标的平均值。

点在多边形内的判别：射线法（铅垂线法、平行线法）、角度算法（弧长法）

射线法：从需判别的点开始划任一方向的直线(该直线可以是铅直线或平行线)，然后计算它所通过多边形的交点。当交点的个数是奇数时，该点位于多边形内；若是偶数，表明它位于多边形外。

角度算法：若点与多边形顶点连线形成的方向角之和为360度——点位于多边形内。否则(等于0度或小于360度)——点位于多边形外。

线与多边形相交的判别：

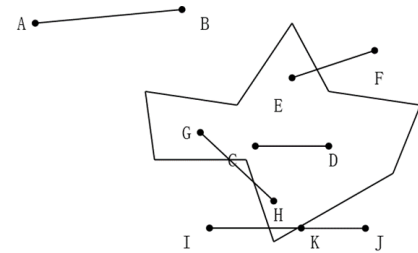
线与多边形是否相交，需要判断每条线段与多边形的边界线段是否有交点。

如果没有任何交点，再判断端点是在多边形内还是多边形外，如果两端点都在多边形外，线段又与多边形不相交，则该线段相离多边形(AB)。

如果两点都在多边形内，并且与多边形边界没有交点，则该线段在多边形内(CD)。

如果有一个或多个交点，该线段与多边形相交，部分在多边形内，部分在多边形外(EF, GH, IJ)。

即使两个点都在多边形内(GH)或都在多边形外(IJ)，它们都可能与多边形相交。



两个多边形是否相交需要判断两个多边形边界的所有线段相互之间是否有交点

- 1) 如果没有任何交点：可能相分离；可能一个多边形在另一个多边形之内
- 2) 两个多边形边界线段只要存在一个交点则表明两个多边形相交；
- 3) 如果它们公共一条边界，则它们相邻

多边形区域填充的算法：递归种子填充算法、扫描线种子填充算法。



➤ 几何变换

1. 几何变换的类型：

- 1) 相似变换：可旋转，保持形状但面积可变。
- 2) 仿射变换：角度可变形，但并行线性性质不变。
- 3) 投影变换：角度、长度、面积均可变形。
- 4) 拓扑变换：拓扑性质，但不保持形状。
- 5) 等积变换，可旋转，保持形状和面积不变。

2. 相似变换：主要用于解决两个坐标系之间的坐标平移或尺度变换

适用条件：①两个坐标系存在夹角；②坐标原点需要平移；③两坐标轴 X, Y 方向具有相同的比例缩放因子。

$$\text{相似变换公式: } \begin{cases} x = m(x' \cos \alpha - y' \sin \alpha) + A_0 \\ y = m(x' \sin \alpha + y' \cos \alpha) + B_0 \end{cases}$$

m ：两坐标系间坐标的比例因子； α ：两坐标系间的坐标轴夹角；

A_0, B_0 ： O' 相对于 O 平移距离；

至少需要 2 对已知的控制点来计算待定系数；超过 2 对坐标，可采用最小二乘法求解

3. 仿射变换：主要用于解决在保留线条平行的前提下，对矩形目标作旋转、平移、倾斜和不均匀缩放等操作。

适用情况：①两个坐标系的原点不同；②两坐标轴在 X, Y 方向的比例因子不一致；

③坐标系之间存在夹角、倾斜等；

$$\text{相似变换公式: } \begin{cases} X = (m_1 \cos \alpha)x - (m_1 \sin \alpha)y + A_0 \\ Y = (m_2 \sin \alpha)x + (m_2 \cos \alpha)y + B_0 \end{cases}$$

m_1 ：X 方向的比例因子 m_2 ：Y 方向的比例因子；

α ：两坐标系之间的坐标轴夹角； A_0, B_0 ： O' 相对于 O 的平移距离

至少需要 3 对已知的控制点来计算待定估算系数。

4. 投影变换：主要用于解决将一种投影坐标转换为另一种投影坐标。

适用条件：数据源与目标数据之间的投影坐标不一致

常用投影方法：解析变换法、数值变换法、数值-解析变换法

- 1) 解析变换法：直接建立两种投影坐标的解析计算公式实现坐标转换（反解变换法、正解变换法、综合变换法）

- A. **反解变换法**：先由原始地图数据的投影坐标 (x, y) 反解出对应的地理坐标 (φ, λ) ，再将地理坐标代入新投影坐标公式，得到 (X, Y)
- B. **正解变换法**：直接建立一种投影坐标到另一种投影坐标的解析关系式： $(x, y) \rightarrow (X, Y)$
- C. **综合变换法**：将上述两种方法进行综合，根据原投影方程反解出纬度(反解)，再根据得到的纬度和 y 值求出两种投影的转换方程式(正解)。
- 2) **数值变换法**：当投影方程未知或变换关系式不易建立时，采用多项式逼近的近似方法建立两种投影之间的变换关系。

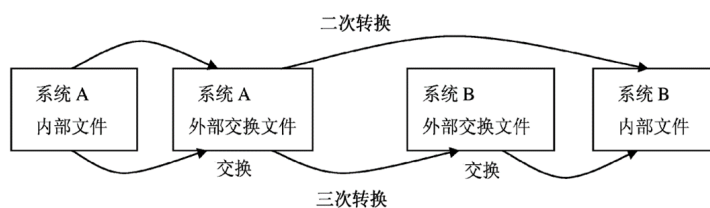
$$\begin{cases} X = a_{00} + a_{10}x + a_{20}x^2 + a_{01}y + a_{11}xy + a_{02}y^2 + a_{30}x^3 + a_{21}x^2y + a_{12}xy^2 + a_{03}y^3 + \dots \\ Y = b_{00} + b_{10}x + b_{20}x^2 + b_{01}y + b_{11}xy + b_{02}y^2 + b_{30}x^3 + b_{21}x^2y + b_{12}xy^2 + b_{03}y^3 + \dots \end{cases}$$

要求得式中的待定系数，**需要选择与待定系数个数相同的同名点**，代入多项式中，求出待定系数，建立投影变换式，为了提高多项式的逼近精度，可将上述**公式取到三次项**。

- 3) **数值-解析变换法**：已知新投影方程式，但不知原投影方程式时，采用多项式方法，将原投影 (x, y) 反解得到地理坐标 (φ, λ) ，再将地理坐标代入新投影方程式，实现两种投影坐标间的转换。

➤ 空间数据转换

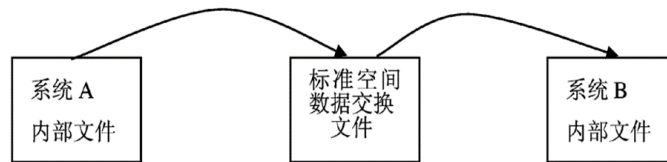
1. **外部交换文件方式**：通过外部数据交换文件进行。



从系统 A 的内部数据转换到系统 B，可能经过 2-3 次转换。

- 1) 先从 A 的内部文件转到 A 的交换文件，如果 B 系统能够直接读 A 系统的交换文件，即转换两次即可；
 - 2) 否则，要从 A 的外部交换文件到 B 的外部交换文件，再从 B 的外部交换文件到 B 的内部文件，此时经过三次转换。
2. **空间数据交换标准方式**：通过空间数据交换标准完成格式转换。

空间数据交换标准：美国 STDS；我国空间数据交换格式 (CNSDTF) 标准

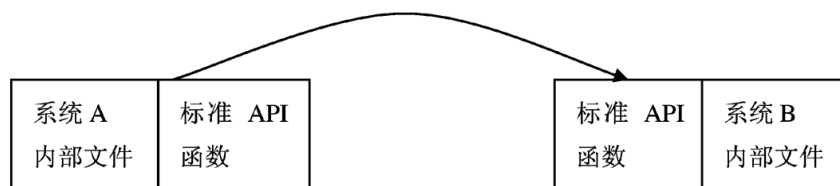


3. 空间数据互操作方式:

基本思想: 制定不同软件的开发标准和接口真正实现异构数据的互操作;

现行标准: 由开放地理信息系统协会(Open GIS Consortium, 简称 OGC)制定的一系列开放标准和接口;

特性: 可扩展性、可升级性、可移植性、开放性、互操作性和易用性;



4. **地理信息 Web 服务:** 以 GIS 相关理论和技术支撑为主要基础, 以计算机网络及其它通信设备等基础设施为操作平台, 融合了 **Web 服务、网络计算**等最新发展成果, 以**满足用户地理信息相关需求为核心的地理信息服务理论和技术系统。**

【不是所有的企业都遵循 OGC 的标准, 其提供的 web 服务不一定是标准的】

Web 服务方式的优点:

- 1) 允许在不同平台上, 以不同语言编写的各种程序, 以基于标准的方式相互通信, 通过 HTTP 协议极大地扩展了传统应用软件地服务范围
- 2) 通过 SOAP、UDDI 和 XML 等标准和技术为应用软件提供了基于 Web 的统一应用标准
- 3) 屏蔽了应用软件底层具体的实现技术

OGC 的 Web 服务标准:

- 1) 矢量数据服务 (Web Feature Service, WFS)

返回的是矢量级的地理标记语言编码

定义的操作:

GetCapabilities、DescribeFeatureType、GetFeature、Transaction、LockFeature

- 2) 栅格数据服务 (Web Coverage Service, WCS)

目的: 面向空间影像数据, 将包含地理位置值的地理空间数据作为栅格在网络中



相互转化

定义的操作: GetCapabilities、GetCoverage、DescribeCoverageType

3) 地图服务 (Web Map Service, WMS)

目标: 利用具有地理空间位置信息的数据制作地图

定义的操作:

A. GetCapabilities: 返回服务级元数据

B. GetMap: 返回地图影像

C. GetFeatureInfo: 返回显示在地图上的某些特殊要素的信息

4) 发表注册服务 (Web Register Service)

➤ 矢量与栅格的数据转换

1. 栅格数据结构:

优点: 具有**属性明显**、**位置隐含**的特点, 易于实现、操作简单, 有利于栅格空间信息模型的分析;

缺点: 数据表达精度不高、数据存储量大。

2. 矢量数据结构:

优点: 具有**位置明显**、**信息隐含**的特点, 操作起来比较复杂, 许多矢量分析操作难于实现;

缺点: 数据表达精度高, 数据存储量小

3. 矢量-栅格转换:

1) **点目标栅格化:** 点之间的简单坐标转换;

2) **线栅格化:** 首先把坐标对变为栅格行列坐标, 然后根据栅格精度要求, 在坐标点之间内插一系列栅格点, 通常利用两点直线方程得到;

3) **面目标栅格化:** 又称为多边形填充, 需要对矢量多边形的边界和内部所有栅格上赋予对应的多边形属性编码, 形成栅格数据阵列, 是矢量栅格化的难点。

4. 面目标栅格化: 边界格元确定+边界格元属性的赋值

5. 边界格元确定主要任务: 确定多边形所对应的栅格单元。

6. 边界格元的确定方法 (5种): ①**内部点扩散算法**、②复数积分算法、③射线算法、④扫描算法、⑤**边界代数算法**。

1) 内部点扩散法:

- A. 按一定栅格尺度将矢量图栅格化后,对矢量图内每个面域多边形分别选择一个内部点(种子点);
- B. 从种子点开始,向其8个相邻栅格扩散,分别判断这8个栅格是否是多边形的边界上:a.若是,则该栅格不作为种子点;b.若不是,则该栅格作为新的种子点;
- C. 新种子点与原种子点一起进行新的扩散;
- D. 重复以上过程,直到所有新老种子点填满该多边形并遇到边界为止。

2) 边界代数算法:

- A. 根据矢量数据的外接矩形,创建目标栅格阵列,并初始化。
- B. 对于任意一个多边形,由边界上某一点开始,逆时针方向搜索边界,下行时,对该线段右侧具有相同行坐标的所有栅格全部减a;平行于栅格行行走时,不作运算。上行时,对该线段左侧具有相同行坐标的所有栅格全部加a;依此处理,直至多边形的所有边处理完。
- C. 重复第二步,直至所有多边形处理完。

7. 边界格元属性的赋值方法:

边界格元属性的赋值: 确定了多边形对应的格网范围后,在同一边界格网单元中可能对应几种不同的属性值,要确定边界格网对应的属性赋值,每一个单元只能取一个值,需要采用合适的取值方法确定其属性取值。

- 1) **中心点法:** 用处于栅格中心处的地物类型或现象特性决定该栅格单元的属性取值。
- 2) **重要性法:** 根据栅格内不同地物的重要性,选取最重要的地物类型决定相应的栅格单元取值。
- 3) **面积占优法:** 以占栅格最大的地物类型或现象特征决定栅格单元的代码。

8. 栅格-矢量数据转换:

- 1) 点的转换: 将栅格点的中心转换为矢量坐标
- 2) 弧段的转换: 提取弧段栅格序列点中心的矢量坐标
- 3) 多边形的转换: 提取以相同编号的栅格集合表示的多边形区域的边界及其拓扑关系,并表示成多个小直线段组成的矢量格式边界线

9. 栅格-矢量数据转换的步骤:



- 1) **多边形边界提取**：将栅格图像二值化或以特殊值标识边界点。
- 2) **边界线追踪**：对每个边界弧段由一个节点向另一个节点搜索。
- 3) **拓扑关系生成**：判断矢量边界与各多边形的空间关系，形成拓扑结构，并建立与属性数据的联系。
- 4) **去除多余点及曲线圆滑**：去除多余点，减少数据冗余；进行光滑处理。
- ◇ **难点**：边界线搜索、拓扑结构生成、多余点去除
10. 栅格-矢量转换算法的分类：边缘跟踪法、边界法、散列线段聚合法等
11. **双边界直接搜索算法的基本思想**：通过边界提取，将左右多边形信息保存在边界点上，每条边界弧段由两个边界链组成，分别记录该边界弧段的左右多边形编号。
12. **边缘跟踪法**：以某一个栅格单元为起点，在该点的八邻域格网单元中按一定方向找出与该栅格单元值相同的栅格单元，重复这个过程直至返回起点。
 - 1) **特点**：①跟踪边界的时候每次都要判断八邻域的值，较费时；②形成拓扑结构较困难；③无法处理单行、单列或单个栅格的情况
 - 2) **改进算法**：
基本思想：①遍历栅格图像，提取结点信息；②追踪结点生成弧段；③连接弧段生成具有拓扑关系的多边形。→基于拓扑关系原理的栅格数据转换矢量方法
13. **基于拓扑关系原理的栅格数据转换矢量方法**：
 - 1) **步骤**：提取结点和坐标点→由结点和坐标点形成弧段→由弧段形成多边形
 - 2) **提取结点和坐标点**：根据相邻4个栅格单元值取值情况，判别找出：
 - A. 结点：只有两种取值的组合，点的走向是唯一的
 - B. 坐标点：三种以上的取值，点的走向有多个

【首尾行的问题：在原始图像的上下各增加一行；左右各增加一列单元值为零的栅格】
 - 3) **由结点和坐标点形成弧段**：找出一个结点作为弧段的始结点，根据结点上连接的弧段的方向，搜索该方向上与之相邻的结点或坐标点；选择始结点上其它未搜索的方向继续进行下一弧段的搜索，直到该结点上连接弧段的方向全部搜索完成；继续下一个结点的搜索工作，如此重复直到结点数据库中所有的结点都完成搜索。
 - A. 坐标点：根据该坐标点的类型和弧段与它连接的方向，决定下一个搜索的方

向;

- B. 结点：将此结点作为该弧段的**终结点**，此弧段的生成工作结束，将该弧段的信息写入弧段数据库，同时在终结点上与该弧段连接的方向作上标记，避免重复搜索。

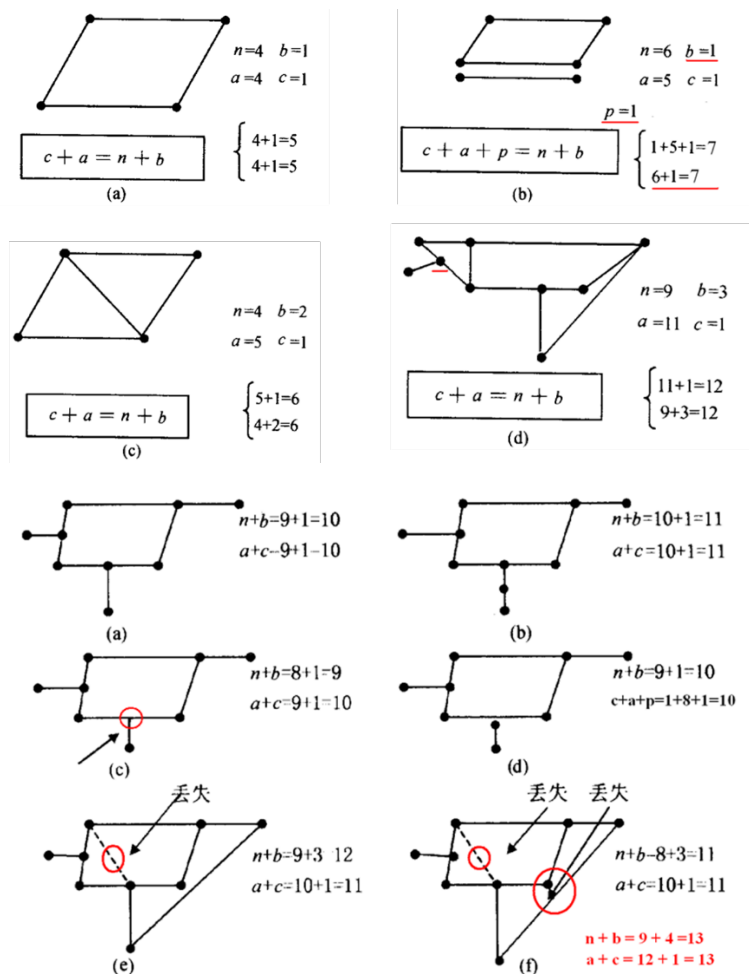
- 4) **由弧段形成多边形**：选择一条弧段作为起始弧段；根据弧段的起始和终结点号、结点类型、弧段的左右多边形属性，搜索与之相连的、属于同一多边形的弧段；如此重复，直到搜索到的弧段与起始弧段连接。

➤ 拓扑关系的自动建立

1. **欧拉定理**：对于一个由若干结点及它们之间的一些不相交的边所组成的图，其**结点数** n 、**弧段数** a 和**多边形数** b 存在下述函数关系：

$$c = n - a + b \text{ 或 } c + a + p = n + b$$

c 是常数，称为多边形地图的特征，如果 b 包含边界里面和外面的面块， c 等于 2，否则 $c=1$ ； p 为独立图形个数减 1，当为一个图形时 $p=0$



2. 欧拉定理是 GIS 拓扑关系检查的基础：

1) 不成立：存在空间关系错误，能发现点、线、面不匹配的情况以及多余和遗漏的图形元素；

2) 成立：不能说明图形空间关系肯定不存在错误。

3. 结点：有 n 结点 (node) 和 v 结点 (vertex) 两种

4. 点-线拓扑关系的建立：

1) 在图形采集与编辑之后，系统自动建立拓扑关系表；

2) 在图形采集和编辑中实时建立，并建立两个文件表分别记录结点所关联的弧段、弧段两端的结点。

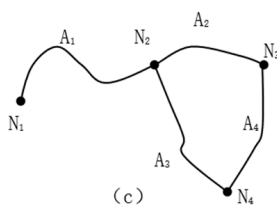


弧段——结点表

Oid	起 点	终结点
A ₁	N ₁	N ₂
A ₂	N ₂	N ₃

结点——弧段表

Oid	弧段号
N ₁	A ₁
N ₂	A ₁ , A ₂
N ₃	A ₂



弧段——结点表

Oid	起 点	终结点
A ₁	N ₁	N ₂
A ₂	N ₂	N ₃
A ₃	N ₂	N ₄
A ₄	N ₄	N ₃

结点——弧段表

N ₁	A ₁
N ₂	A ₁ , A ₂ , A ₃
N ₃	A ₂ , A ₄
N ₄	A ₃ , A ₄

5. 多边形的种类：

1) 独立多边形：与其它多边形没有共享边界。

2) 有公共边的简单多边形：仅采集弧段数据，然后将多边形的边界聚合起来，建立多边形文件。

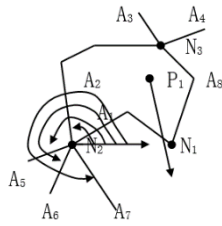
3) 带岛屿的多边形：除了按第二种方法建立多边形，还要考虑内岛。

4) 复合多边形：由两个或多个不相邻的多边形组成，对这种多边形一般是在建立单个多边形以后，再用人工或某一种规则组合成复合多边形。

6. 基于方位角的多边形拓扑关系自动建立方法：

本质：计算得到多边形-弧段关系表和弧段-多边形关系表。

基本步骤：确定弧段邻接关系→计算弧段方位角→搜索多边形→确定拓扑关系。



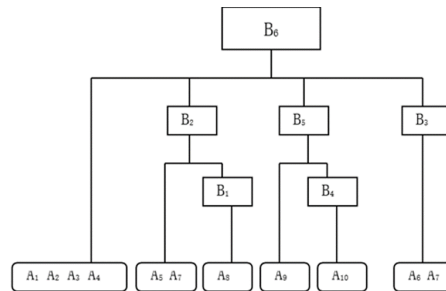
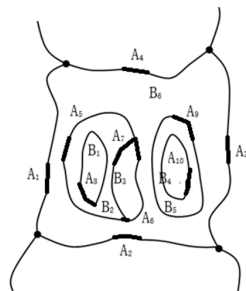
多边形——弧段关系

oid	弧段号
P ₁	A ₁ A ₂ A ₈ ,

弧段——多边形关系

Oid	左面	右面
A ₁	P ₁	
A ₂	P ₁	
⋮		
A ₈		P ₁

7. 带岛屿多边形拓扑关系的建立：先建立简单多边形，再采用两个多边形相交判断方法



判断一个多边形包含那些多边形，并采用树形结构描述它们之间的关系。

➤ 矢量数据错误与编辑

1. 矢量数据错误类型：

1) 定位错误：几何要素存在的几何位置错误。

✧ 主要表现为几何要素的缺失或者位置错误

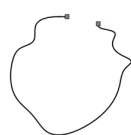
✧ 定位错误与所使用的数据源密切相关

A. 二手数据源：如已有的纸质地图，定位错误主要由数字化地图与原始纸质地图的匹配程度导致；

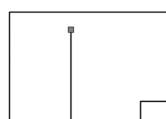
B. 一手数据源：如 GNSS 或遥感影像等，其数字化精度主要受到数据源自身的精度所限制，如 GNSS 误差和遥感影像分辨率等。

2) 拓扑错误：空间要素之间存在空间逻辑不一致。

A. 面状（多边形）目标错误：未闭合的多边形、两个相邻的多边形间存在缝隙多边形重叠；



(a) 未闭合



(b) 多边形间存在裂隙



(c) 多边形重叠。

B. 线状目标错误:

- a) 欠头: 线间存在缝隙, 称为欠头 (undershoot);
- b) 过伸: 若一条弧段过长称为过伸 (overshoot),

【这两类错误都表现为存在悬挂结点】

- c) 线段方向也可能存在拓扑错误

C. 点状目标错误: 一个多边形存在多个内部标识点等

D. 不同类型空间目标间错误: 两个或两个以上图层中不同类型空间要素的拓扑错误。

2. 拓扑错误的检查方法: 计算机自动拓扑数据检查

检查内容: 结点是否匹配; 是否存在悬挂线; 多边形是否闭合;

3. 几何数据编辑方法:

1) 结点的编辑: 建立点线面关联拓扑关系的桥梁和纽带。

A. 结点吻合 (Snap): 即结点匹配和结点附合

结点匹配方法:

- a) **结点移动**: 将 B 点和 C 点移到 A 点;
- b) **用鼠标拉一个矩形**, 落入在这矩形内的结点坐标符合成一致, 即求它们的中点坐标, 并建立它们之间的关系;
- c) **求交点**: 求两条线的交点或延长线的交点, 即是吻合的结点;
- d) **自动匹配**: 给定一个容差, 在容差范围之内的结点自动吻合在一起。

B. 结点与线的吻合: 结点可能不完全交于线目标上, 需要进行编辑, 进行结点与线的吻合

结点匹配方法:

- a) **结点移动**: 将结点移动到线目标上;
- b) **使用线段求交**: 求出 AB 与 CD 的交点;
- c) **使用自动编辑的方法**: 在给定的容差内, 利用自动求交实现吻合。

【会吻合到**线的结点**上去】

C. 伪结点: 有且仅有两个线目标相关联的结点。【可清除可不清除】

2) 图形编辑:

A. 删除/增加顶点:

- a) 删除顶点：删除顶点后，线目标的节点个数比原来少，线目标不用整体删除，只是重写一次坐标，拓扑关系不变。
- b) 增加顶点：找到增加顶点对应的线段，定义一个新顶点位置。
- B. 移动顶点：只改变某个点的坐标，不涉及拓扑关系的维护和调整。通过移动顶点，可以改变弧段的形状，使其更好地拟合原有地物边界。
- C. 删除弧段：先要把原来的弧段分成三段，（在存贮上原来的弧段实际上被删除），去掉中间一段，保留两端的两条弧段。由于赋了两个新的目标标识，原来建立的空间拓扑关系，需要进行调整 and 改变

➤ 图幅拼接与接边

1. 逻辑裂隙与几何裂隙：

- 1) 逻辑裂隙：当一个地物在一幅图的数据文件中具有地物编码 A，而在另一幅图的数据文件中却具有地物编码 B，或者同一个物体在这两个数据文件中具有不同的属性信息，如公路的宽度，等高线的高程等。
- 2) 几何裂隙：一个地物的两部分不能精确地衔接。接边：在拼接过程中，把相邻图幅的空间数据在逻辑上和几何上融成一个连续一致的数据体。

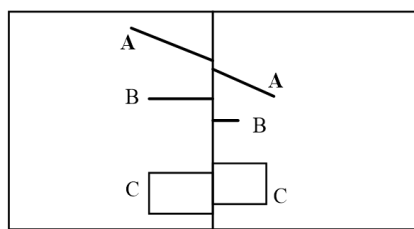
2. 接边：在拼接过程中，把相邻图幅的空间数据在逻辑上和几何上融成一个连续一致的数据体。

3. 几何接边：

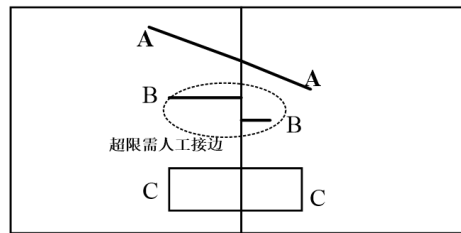
- 1) 步骤：调出需要接边的两幅或多幅图数据 → 以其中的一个作为活动图幅（或称活动工作区），其它图幅的作为参考，沿图幅的边缘选取一定范围的空间目标。
→ 以活动工作区的目标为基准，根据图廓边上弧段的结点坐标查找相邻图幅对应弧段。
- 2) 自动吻合的条件：地物编码相同 + 结点坐标在一定的容差范围内
【将两边结点坐标取中数，空间关系不变】



- 3) 人工接边的情况：地物编码不同 or 超过接边的匹配容差。



接边前



自动接边后

4. 逻辑接边：

- 1) 步骤：检查同一目标在相邻图幅的地物编码和属性赋值是否一致 → 将相邻图幅的同一空间实体目标数据连在一起。
- 2) 两种方法：
 - A. 建立一个新的文件，新文件里存储该目标的所有位置信息；
 - B. 不建立新文件，仅利用关键字，系统自动在周边图幅的文件中搜索同一目标

第六章 · 地理空间数据处理

➤ 数据管理概述

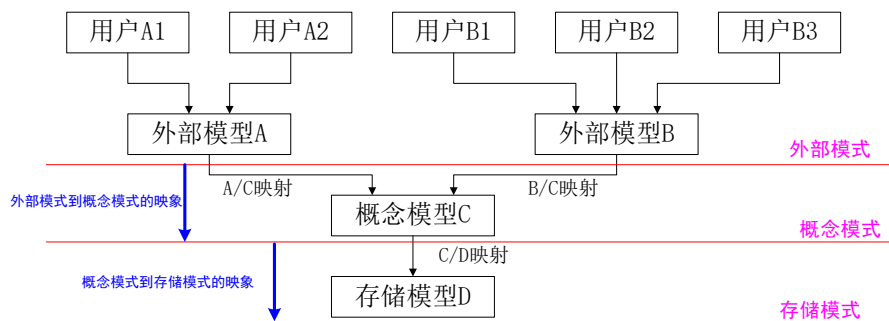
1. **数据管理**：利用计算机软硬件技术，对收集的数据以数据文件或者数据库的形式存储在计算机中，并进行查询、处理及应用的过程。
2. **数据组织的层次**可以有两类分级方法：**按逻辑单位分级、按物理单位分级**：
 - 1) 数据物理单位是指**数据在存储介质上的存储单位**，层次结构为：比特、字节、字、块（物理纪录）、桶和卷；
 - 2) 数据逻辑是**按数据与所描述的对象关系**分，单位的层次有：数据项、记录、文件、数据库。
3. **数据库中数据的组织分级**：
 - 1) 数据项（基本项、字段）：定义数据的最小单位；数据项的取值范围称为域；
 - 2) 记录：由若干相关联的数据项值组成；为了唯一标识每一条记录，每条记录必须有一个标识符，也叫关键字；
 - 3) 文件：给定类型的记录的全部具体值集合；
 - 4) 数据库：具有特定联系的数据集合，内部表现为文件的集合



4. **数据间的逻辑关系**：记录与记录之间的关系。
包括：一对一的关系、一对多的关系、多对多的关系
5. **文件组织形式**：顺序文件、索引文件、直接文件、倒排文件
 - 1) 顺序文件：**按记录的主关键字排序**；
在物理存储上可以有三种形式：向量方式、链方式、块方式
 - A. 向量方式：物理结构与逻辑结构一致；查找方便；插入记录困难
 - B. 链方式：文件的逻辑顺序靠链来实现；每条记录含有一个指针（下一记录）
 - C. 块方式：把文件分为块；块之间用指针连接；块内按顺序存储的
 - 2) 直接文件：由关键字直接映射存储位置，随机查询；
☆可能存在“冲突”问题：可能会存在多个记录的存储地址相同的问题
 - 3) 索引文件：添加索引表文件；
 - 4) 倒排文件：**按次关键字**来添加辅助索引
6. **数据管理方式**：
 - 1) 程序管理：数据和程序文件混合存储；
 - 2) 文件管理：数据和程序文件分开，由文件系统管理数据。
☆特点：**数据共享性差、冗余度大、数据关联性差、不一致的可能性大、安全性问题**
 - 3) 数据库管理：由数据库方式管理数据
7. **数据库的主要特征**：
 - 1) 数据集中控制：集中控制和控制有关数据，以保证用户和应用可以共享数据
 - 2) 数据冗余度小：消除冗余数据，但有时适当冗余可以提高数据处理效率
 - 3) 数据独立：数据与应用程序相互独立
 - 4) 复杂数据模型：与文件方式本质区别，复杂数据模型是实现数据集中管理、减少数据冗余的前提
 - 5) 数据保护：为数据提供保护措施，包括：安全性控制（权限）、完整性控制（数据正确、有效、相容）、并发控制、故障的发现和恢复
8. **GIS数据库的建库要求**：遵循通用数据库的原理和方法 + 考虑地理空间数据的特点。
9. **GIS数据库与一般数据库的区别**：
具有空间数据和属性数据；数据量大；数据具有非结构化的特点；数据应用面广
10. **数据库的基本结构是三级模式结构**：

外模式（用户模式）、模式（概念模式）和内模式（存储模式）

- 1) 存储模式（Internal Schema）是数据物理结构和存储方式的描述，是数据在数据库内部的表示方式
- 2) 概念模式（Schema）是数据库中全体数据的逻辑结构和特征的描述，是所有用户的公共数据视图；它是数据库系统模式结构的中间层，既不涉及数据的物理存储细节和硬件环境，也与具体的应用程序、开发工具无关
- 3) 用户模式（External Schema）是数据库用户能够看见和使用的局部数据的逻辑结构和特征的描述，是数据库用户的数据视图，是与某一应用有关的数据的逻辑表示；用户模式是概念模式的子集；用户模式是保证数据库安全性的一个有力措施



11. 数据库系统的系统构成（4个部分）：

数据库；数据库管理系统（及其开发工具）；应用系统；数据库管理员和用户

12. 数据库管理系统（Database Management System，简称DBMS）是处理数据库存取和各种管理控制的软件，它位于用户与操作系统之间的一层数据管理软件，提供数据库的用户接口。【提供一个可以方便地、有效地存取数据库信息的环境】

- 1) DBMS 的组成：数据、元数据；存储管理程序；事务管理程序；查询处理程序
- 2) DBMS 的主要功能包括：
 - A. 数据定义功能：数据定义语言（Data Definition Language，简称DDL）
 - B. 数据操纵功能：数据操纵语言（Data Manipulation Language，简称DML）
 - C. 数据库的运行管理：在系统运行过程中保障数据的安全性、完整性、一致性，并发使用，故障恢复
 - D. 数据库的建立和维护功能

13. 数据库模型：层次模型、网络模型、关系模型、面向对象模型

- 1) 层次模型是数据库系统中最早出现的数据模型，层次模型是一种树结构模型。

把满足以下两个条件的基本层次联系集合称为层次模型：



①有且只有一个结点没有双亲结点，这个结点称为根结点；②根以外的其他结点有且只有一个双亲结点；

✧ 层次模型不能表示多对多的关系。

2) 网络模型：在数据库中，把满足以下两个条件的基本层次联系集合称为网络模型

①允许一个以上的结点无双亲；②一个结点可以有多个的双亲

✧ 结点之间没有明确的从属关系，一个结点可与其他多个结点建立联系（是多对多的关系）；消除数据冗余；支持拓扑表达；

3) 关系数据模型：数据的逻辑结构归结为满足一定条件的二维表，表格之间通过关键字关联。

✧ 结构灵活、易于理解、维护方便；地理对象语义薄弱

4) 面向对象模型：对象和类是核心。

✧ 具有可扩充性；可以模拟和操纵复杂对象

➤ 空间数据组织

1. 空间数据组织：指按照一定的方式和规则对空间数据进行整理与存储的过程

2. 图幅内空间数据的组织：工作区、逻辑层、地物类

1) 工作区：一幅图或几幅图的范围，工作区范围内包含了所有层的空间数据

2) 逻辑层：便于出图。如果一个工作区包含的内容很多，例如所有地物类，这时为了显示、制图和查询方便，需要定义逻辑层。

3) 地物类：类型相同的地物组合，一个工作区通常包含多个地物类。

☆不同几何类型的地物（点状、线状、面状）必须分不同的数据文件

3. 图库管理：GIS 大型工程管理的一种模式，建立图幅索引，即基于工作区的二维空间索引。

4. 数据库组织方式：矢量数据组织、栅格数据组织、无缝数据库

1) 矢量数据组织：

A. 主要内容：空间点、线、面等要素的表达、组织和存储

B. 组织方式：

空间实体的几何形状：坐标；

空间实体的属性特征：属性表；

空间实体的空间关系：建立拓扑关系

- C. **对象-关系存储方案**：几何数据以自定义格式的二进制形式和其相关属性数据一起存储在关系表中；充分利用关系数据库的优点进行数据管理查询；关系表的组织方式依数据库软件不同而有所差异

✧ 不保存对象之间的拓扑关系；空间属性数据采用关系模型进行组织；空间几何数据使用对象进行表达；对象多采用较为 OGC 简单要素模型

- D. **OGC 简单要素模型**：定义了一组与平台无关的空间二维几何对象模型和一系列函数访问接口。

几何对象模型：对地理几何类型和空间参考的定义：**支持 WKT、WKB**

【数据格式：用于空间要素的交换与存储】

- a) Well-Known Text (WKT)：一种文本标记语言，表示空间几何对象及空间参考信息
- b) Well-Known Binary (WKB)：WKT 的二进制形式，在数据传输与数据库存储时用

函数访问接口：**支持 SQL 中的实现**

2) 栅格数据组织：

- A. **金字塔层—波段—数据分块多级组织机制**：

影像金字塔：指在**同一的空间参照**下，根据用户需要以**不同分辨率**进行存储与显示，形成分辨率由粗到细、数据量由小到大的**多层级影像数据结构**

- B. 大部分对象-关系数据库系统支持栅格数据存储：GeoRaster、PostGIS

3) 无缝数据库：

组合方式：

- A. 物理无缝：不分幅，将某区域内的地理数据统一存储在数据库中，实现空间数据库的物理无缝连接；
- B. 逻辑无缝：在大型 GIS 图库工程管理中，选择物理分幅，但保持逻辑上空间数据的无缝。

➤ 空间数据索引

1. **空间数据索引**：依据空间对象的位置和形状或空间对象之间的某种空间关系按一定的顺序排列的一种数据结构，是实现空间数据快速查询和检索的重要手段。
2. **空间数据索引的种类**：对象范围索引、空间格网索引、二叉树空间索引、四叉树空间

索引、R 与 R+树空间索引。

3. **对象范围索引**：当记录每个空间对象的坐标时，同时记录其对象的最大、最小窗坐标。根据空间对象的最大最小范围，预先排除那些没有落入检索窗口内的空间对象。
 - ◇ 窗坐标：完全包含对象的最小外接矩形对角点坐标
4. **空间格网索引**：将工作区按一定的规则划分为格网，记录每一格网内所包含的空间对象，建立空间索引的线性表，将格网按 Morton 码进行编码，建立 Morton 码与空间对象的关系。
5. **二叉树空间索引**：二叉树是每个节点最多有两个子节点的树结构。常见的基于二叉树的空间索引结构有 BSP（二叉空间分割，Binary Space Partitioning）树，KD（k-dimensional tree）树
 - 1) BSP 树采用二叉树索引思想将目标空间逐级一分为二的划分，**BSP 树能很好地与空间数据库中空间对象的分布情况相适应**
 - 2) KD 树是 BSP 树向多维空间的扩展，用 $k-1$ 维超平面把 k 维空间递归地分割成两部分，是对数据点在 k 维空间中划分的一种数据结构，**主要应用于多维空间关键数据的搜索**
6. **四叉树空间索引**：分为线性四叉树空间索引和层次四叉树空间索引
 - 1) 线性四叉树空间索引：先采用 Morton 编码（Peano 键），再根据空间对象覆盖的范围，进行四叉树分割。**【只存储最后叶节点信息】**
 - 2) 层次四叉树空间索引：需要记录中间节点和父节点到子节点之间的指针
7. **R 树**：设计一些虚拟的矩形，将空间对象包含在这些矩形框内，矩形框包含空间对象的指针，判断哪些虚拟矩形落在检索窗口内，进一步判断虚拟矩形中的哪些空间对象为被检索对象。
8. **基本 R 树构造矩形框原则**：矩形之间尽可能少重叠；矩形尽可能包含更多的空间对象；矩形可以嵌套，即矩形可以包含更小的矩形；一个空间对象通常仅被一个虚拟矩形所包含。
9. **R 树存在的问题**：R 树要求虚拟矩形一般尽可能少地重叠，并且一个空间对象通常仅被一个虚拟矩形所包含，但空间对象的最小矩形范围经常重叠→R+ 树（改进的 R 树空间索引）
10. **R+ 树（改进的 R 树空间索引）**：允许虚拟矩形相互重叠；允许一个空间目标被多个虚

拟矩形所包含（将重叠部分切割成多个部分）

➤ 空间数据库管理系统

1. 文件与关系数据库混合管理

第一阶段：两个子系统分别存储和检索空间数据与属性数据。

属性数据存储在售用的 RDBMS 中，几何数据存储在文件系统中，两个子系统间用标识符联系起来（关键字）。

要同时启动两个系统（GIS 图形系统和关系数据库管理系统），甚至两个系统来回切换，使用起来很不方便。

第二阶段（混合处理模式）：利用数据库系统提供的与高级语言(C, Fortran)的接口，如 SQL 语句(通过 C 语言)、ODBC(开放性数据库连接协议)，使得 GIS 在 C 环境下可以操纵数据库。

用户只需要开发 GIS 与 ODBC 的接口软件，就可以将属性数据与任何一种支持 ODBC 协议的关系数据库管理系统连接，不需要启动一个完整的数据库管理系统，图形和属性数据的查询与维护完全在一个界面下。

混合管理模式的特点：

- 1) 由于空间数据和属性数据分开存储，在表现地理空间数据方面缺乏完整的表达语义和存储机制
- 2) 难于保证数据存储、操作的统一
- 3) 系统切换、使用不方便
- 4) 在数据的安全性、一致性、完整性、并发控制以及数据损坏后的恢复方面缺少基本功能

2. 全关系型空间数据库管理系统

图形数据和属性数据都采用关系数据管理系统管理，即使用 **同一 DBMS 管理**

在标准的关系数据库上增加 **空间数据管理层**

利用该层将结构查询语言（GeoSQL）转化成标准的 SQL 查询

借助索引数据的辅助关系实施空间索引操作

模式1（基于关系模型方式）：图形数据都按照关系数据模型组织，涉及一系列关系连接运算，相当费时，处理空间目标效率不高

模式2: Blob 字段

基本思想：

将图形数据的变长部分处理成二进制块 BLOB 字段

大部分关系数据库管理系统都提供了二进制块的字段域

特点：

省去了前面所述的大量关系连接操作

二进制块的读写效率要比定长的属性字段慢得多，特别是涉及对象的嵌套时，速度更慢

3. 空间数据库引擎 SDE (Spatial Database Engine):

建立在现有关系数据库基础上

介于 GIS 应用程序和空间数据库间的中间件，为用户提供访问空间数据库的统一接口

SDE 引擎本身不具有存储功能，只提供和底层存储数据库之间访问的标准接口

SDE 屏蔽了不同底层数据库的差异，建立了上层抽象数据模型到底层数据库之间的数据映射关系

空间数据库引擎支持矢量和栅格数据在关系数据库中的存储

4. 对象关系数据库系统

对传统的关系数据库进行扩展，使之能够直接存储和管理非结构化的空间数据

空间扩展完全包含在 DBMS 中

特点：

- 1) 解决了空间数据的**变长记录管理问题**，由数据库软件商扩展，效率比自行采用二进制块的方式高；
- 2) 对空间对象的数据结构进行预先的定义，用户不能再定义，使用受到限制；
- 3) 对象之间一般不能带有拓扑关系，没有解决对象嵌套的问题，因此**具有拓扑数据结构的数据不能采用对象-关系数据库进行管理。**

5. 面向对象空间数据库管理系统

➤ 空间数据查询

1. **空间查询**：从数据库中找出符合该条件的空间数据的过程。

☆空间查询是空间分析的基础，任何空间分析始于空间查询

2. **空间查询的分类**：

- 1) **基于属性特征的查询**：根据空间对象或实体的**属性数据**来查询满足给定条件的地物或区域的空间位置，统计其几何与属性参数的查询方式

- 2) **基于空间位置的查询**：指根据空间对象或实体的地理位置查询其满足条件的实体集合及其属性信息的查询

按点查询、按矩形查询、按圆查询、按多边形查询

- 3) **基于空间关系的查询**：包括空间拓扑关系查询和缓冲区查询。空间关系查询有些是通过拓扑数据结构直接查询得到，有些是通过空间运算，特别是空间位置的关系运算得到

邻接查询、包含关系查询、穿越查询、落入查询、缓冲区查询

- A. 邻接查询：

- a) 第一类是多边形邻接查询，如查询与面状地物 A 相邻的所有多边形。

- b) 第二类是线与线的邻接查询，例如查询所有与主河流 A 关联的支流

- B. 包含查询：查询某一个面状地物所包含的某一类的空间对象。这种查询使用空间运算执行。

- C. 穿越查询：使用空间运算执行。根据一个线状目标的空间坐标，计算出哪些面状地物或线状地物与它相交。

- D. 落入查询：执行这一操作采用空间运算即可，即使用点在多边形内，线在多边形内，或面在多边形内的判别方法。

- E. 缓冲区查询：缓冲区查询不对原有图形进行切割，只是根据用户需要给定一个点缓冲，线缓冲或面缓冲的距离，从而形成一个缓冲区的多边形，再根据前面所述的多边形检索的原理，检索出该缓冲区多边形内的空间地物

3. **SQL 语言**：结构化查询语言 SQL (Structured Query language) 是一种数据库查询和程序设计语言，用于存取数据以及查询、更新和管理关系数据库系统。

SQL 主要提供了**数据定义**、**数据操作**、**数据查询**和**数据控制**等几个部分的功能

可以把 SQL 分为两个部分：数据操作语言 (DML) 和 数据定义语言 (DDL)。

DML:

SELECT - 从数据库表中获取数据

UPDATE - 更新数据库表中的数据

DELETE - 从数据库表中删除数据

INSERT INTO - 向数据库表中插入数据

DDL: 用于创建或删除表格

CREATE DATABASE - 创建新数据库 ALTER DATABASE - 修改数据库

CREATE TABLE - 创建新表 ALTER TABLE - 变更（改变）数据库表

DROP TABLE - 删除表 CREATE INDEX - 创建索引（搜索键）

DROP INDEX - 删除索引

拓展的 SQL 语言：扩展的空间 SQL 查询语言主要通过引入空间数据类型和空间操作

算子来扩展 SQL，使得用户能够使用 SQL 语法进行空间对象的查询

常用的空间关系谓词

Ajacent—相邻 Contain—包含 Cross—穿过 Inside—在内……之内 Buffer—缓冲区

➤ 新型空间数据库系统

1. Not Only SQL:

1) NoSQL 存储:

- a) 不使用关系数据模型
- b) 数据模型多种多样，不需要事先为存储数据建立模式，可随时存储自定义的数据格式
- c) 能很好地处理半结构和非结构化的大数据

2) NoSQL 特点:

- a) 针对大型集群而设计
- b) 支持数据自动分片
- c) 很容易实现水平扩展
- d) 具有良好的伸缩性

	关系数据库	NoSQL数据库
关系模型	是	否
SQL查询语言	是	否
事务ACID特性	是	否
水平扩展	否	是
海量数据	否	是
动态模式	否	是

3) NoSQL 数据系统分类:

- a) 列簇存储：以列为单位存放数据，列的集合称为列簇，每一列中的每个数据项包含时间戳属性，可保存列中的同一个数据项的多个版。【谷歌的 BigTable，后期衍生的 HBase，Cassandra 等】
- b) 键值模型：把“键”映射到相应的“值”（数据），而不关心数据的内容，非结构化的数据存储模式，应用开发者需要自己定义、解析“值”的数据格式，不支持任何非“键”的查询。【Amazon Dynamo】
- c) 文档模型：将数据封装存储在 JSON 或者 XML 等类型的文档中
- d) 图模型：基于图结构，使用节点、关系、属性三个基本要素存放数据之间的

存储模型	代表性数据库产品	特点	应用场景
列簇存储	Google BigTable, Apache HBase, Cassandra	快速数据聚合, 可扩展性强, 版本控制, 分布式存储	大量数据的高访问负载, 日志系统等
键值模型	Amazon Dynamo, Riak, Redis, Oracle Berkeley DB	模型简单, 操作简单, 查找速度快, 分布式存储	内存缓存, 网站分析, 电子商务等
文档模型	MongoDB, CouchDB, CouchBase	非结构化文档存储, 面向对象	内容管理系统, 博客系统, 事件日志等
图模型	Neo4j, Titan, OrientDB	数据关联性, 严格的数据模式	社交网络, 推荐系统等

关系信息

4) NoSQL 数据系统分类：

- a) 优点：具有高扩展性、支持分布式存储、高性能，架构灵活、支持结构化，半结构化以及非结构化数据、运营成本低。
- b) 缺点：没有标准化的数据模型和查询语言、查询功能有限、大部分不支持数据库事务。

2. 关系数据库与 NoSQL 数据库的对比：关系数据库使用关系模型，使用 SQL 查询语言，遵从事物 ACID 特性，不支持水平扩展，不适用海量数据，不支持动态模式。NoSQL 数据库不使用关系模型，不使用 SQL 查询语言，不遵从事物 ACID 特性，支持水平扩展，适用海量数据，支持动态模式。

3. 对于空间数据来说，哪种 NoSQL 数据库使用广泛(更适用于空间数据管理)?

答：MongoDB 和 Neo4j。因为 MongoDB 和 Neo4j 给出了空间拓展

- 1) MongoDB 的空间拓展：一种文档数据库，提供高性能，高可用性和易可伸缩性。文档可以包含许多不同的键 - 值对，或键 - 数组对，或者是嵌套的文档。

MongoDB 将数据存储在 JSON 文档中。不同于传统关系型数据库，MongoDB 采

取了动态模式设计：相同集合中的文档不需要具有相同的字段或结构，且集合中文档的公共字段可以保存不同类型的数据。

- 2) Neo4j 的空间拓展：基本思想是将空间数据映射为图模型，空间对象存储为图的节点，对象间的关系存储为图的边。空间实体的几何属性可以使用标准的 WKT 或 WKB 格式保存为图节点的属性。空间数据用 CQL 语言查询，使用 R 树索引。集成了开源的地理空间几何库 Java Topology Suite 可进行常用的空间分析。

第七章 · 空间数据分析

➤ 矢量数据空间分析

1. 叠置分析：

叠置分析是 GIS 中的一项重要非常重要的空间分析功能。

空间叠置至少涉及到两个图层：

- 1) 至少有一个图层是多边形图层，称为**基本图层**；
- 2) 另一图层可能是点、线或多边形，称为**上覆图层**。

2. 空间叠置分析往往涉及到**逻辑交**、**逻辑并**、**逻辑差**的运算

3. 矢量数据叠置分析：

- 1) **点与多边形的叠置**：将一个含有点的图层叠加上另一个含有多边形的图层上，以**确定每个点落在哪个多边形内**。通过点在多边形内的判别完成的，得到一张**新的属性表**，该属性表除了原有的属性以外，还含有落在那个多边形的目标标识。
- 2) **线与多边形的叠置**：将**线的图层叠置在多边形的图层上**，以确定一条线落在哪一个多边形内。往往一个线目标跨越多个多边形，这时需要先进行**线与多边形边界的求交**，并将**线目标进行切割**，形成一个新的空间目标的结果集。建立起**线状目标的属性表**：包含原来线状目标的属性和被叠置的面状目标的属性。
- 3) **多边形与多边形的叠置**：需要将两层多边形的边界**全部进行边界求交的运算和切割**。然后根据切割的弧段**重建拓扑关系**，最后判断新叠置的多边形分别落在原始多边形层的哪个多边形内，建立起叠置多边形与原多边形的关系。

【设两个原始的多边形图层一个称之为**本底多边形**，另一个称之为**上覆多边形**，叠置得到的新多边形称为**叠置多边形**。】

✧ 据叠置分析过程中逻辑运算的不同，可以得到不同的分析结果

4. **矢量叠置与栅格叠置的区别**：栅格数据叠置的直观概念就是将两幅或多幅地图重迭在一起，产生新多边形和新多边形范围内的属性。其结果制虽然数据存储量小，但是运算过程复杂。矢量数据在叠置地图的相应位置上产生新的属性的分析方法。其结果虽道然数据存储量大，但是运算过程较简单。
5. **缓冲区分析**：是对一组或一类地物按缓冲的距离条件，建立缓冲区多边形图层，然后将这一个图层与需要进行缓冲区分析的图层进行叠置分析，得到所需要的结果。

分类：点缓冲区、线缓冲区、面缓冲区、栅格缓冲区（经过距离变换，计算出背景像元与空间目标的最小距离）

两步操作：

- 1) 第一步是建立缓冲区图层；
- 2) 第二步是进行叠置分析。

缓冲区的建立方法：建立点缓冲区仅是以点状地物为圆心，以缓冲区距离为半径绘图即可；线状地物和面状地物的缓冲区的建立也是以线状地物或面状地物的边线为参考线，作它们的平行线，再考虑端点圆弧，即可建立缓冲区。

◇ 有时需要根据空间地物的特性不同，建立不同距离的缓冲区

在实际处理中，按照常规算法建立的缓冲区，缓冲区之间往往出现重叠，缓冲区可能彼此相交。

方法 1：在作参考线的平行线时，考虑各种情况，自动切断彼此相交的弧段。

方法 2：通过叠置的缓冲区多边形进行合并，并清除缓冲区内的相交弧段。

6. **网络分析**：全面地描述由**点和线组成的网状事物**以及其间的相互关系和内在联系就必须利用基于此类数据所进行的一类空间分析——网络分析。

在数学领域内，网络分析的基础是**图论和运筹学**，它通过研究网络的状态以及模拟和分析资源在网络上的流动和分配情况，对网络结构及其资源等的优化问题进行研究。



包括：最佳路径、资源分配、结点或弧段的游历（旅行推销员问题、中国邮递员问题）以及最小连通树、最大（小）流等问题。

在 GIS 中，网络分析是依据网络拓扑关系（线性实体之间、线性实体与结点之间、结点与结点之间的连接、连通关系），通过考查网络元素的空间及属性数据，以数学理论模型为基础，对网络的特征和性能进行多方面的分析计算。

1) **路径分析**：路径分析是用于模拟两个或两个以上地点之间资源流动的路径寻找过程。路径分析是 GIS 中最基本的功能，其核心是对最佳路径的求解。在指定网络的两结点间找一条阻抗强度最小的路径。其求解方法有几十种，而 Dijkstra 算法被 GIS 广泛采用。

较好的近似解法有：

基于贪心策略的最近点连接法、最优插入法、基于启发式搜索策略的分枝算法、基于局部搜索策略的对边交换调整法等。

2) **资源分配（也称定位与分配问题）**：①目标选址②将需求按最近原则寻找供应中心（资源发散或汇集地）。常用的算法是 P 中心模型

3) **连通分析**：

- a) 连通分量求解：从某一结点或边出发能够到达的全部结点或边。
- b) 最少费用连通方案的求解：在耗费最小的情况下使得全部结点相互连通。
- c) 连通分析对应图的生成树求解，通常采用深度优先遍历或广度优先遍历生成相应的树。
- d) 最少费用求解过程：是生成最优生成树的过程。

4) **流分析**：流就是资源在结点间的传输，流分析的问题：按照某种优化标准（时间最少、费用最低、路程最短或运送量最大等）设计资源的运送方案。

7. 网络分析之最佳路径问题：

最佳路径中的佳包含很多含义：可以指一般地理意义上的距离最短，还可以是时间最短、费用最少、线路利用率最高等标准；但是无论引申为何种判断标准，其核心实现方法都是最短路径算法。

路径分析的分类:

- 1) 静态求最佳路径: 在给定每条链上的属性后, 求最佳路径;
- 2) N 条最佳路径分析: 确定起点或终点, 求代价最小的 N 条路径, 因为在实践中最佳路径的选择只是理想情况, 由于种种因素而要选择近似最优路径;
- 3) 最短路径或最低耗费路径: 确定起点、终点和要经过的中间点、中间连线, 求最短路径或最小耗费路径;
- 4) 动态最佳路径分析: 实际网络中权值是随权值关系式变化的, 可能还会临时出现一些障碍点, 需要动态的计算最佳路径 (临时改变障碍)。

求最短路径的基本思路:

- 1) 设欧氏空间的任意三点 x, y, z , 令 $d(x, y)$ 为 $x \rightarrow y$ 的距离, 则有:

$$d(x, y) \leq d(x, z) + d(z, y)$$

- 2) 令 d_k 为结点 v_i (v_1) 到 v_j 的最短距离, w_{ij} 为 v_i 到 v_j 的权值, 对于 (v_i, v_j) 不属于边的结点对时, 令 $w_{ij} = \infty$, 显然

$$d_1 = 0 ; d_k \leq d_j + w_{jk} , k, j = 2, 3, \dots, p$$

- 3) 当且仅当 (v_j, v_k) 在 v_1 到 v_k 的最短路径上时, 等式成立; 假设 d_k 是 v_1 到 v_k 的最短路径、该路径最后一弧段为 (v_j, v_k) , 由局部与整体的关系, 路径的前一段 v_1 到 v_j 也必为 v_1 到 v_j 的最短路径, 则

$$d_1 = 0 ; d_k = \min(d_j + w_{jk}) , k, j = 2, 3, \dots, p ; k \neq j$$

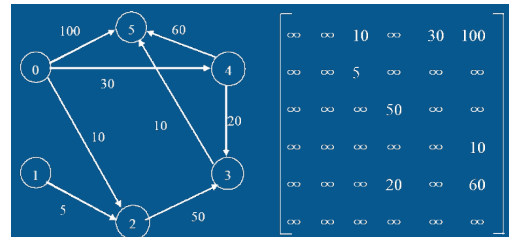
- 4) 上式是最短路径方程,

计算最短路径的 Dijkstra(迪杰斯特拉)算法

Dijkstra 算法可以用于计算从有向图中任意一个节点到其它节点的最短路径。

Dijkstra(迪杰斯特拉)算法步骤:

- 1) 用带权的邻接矩阵 $Cost$ 来表示带权的 n 个节点的有向图, $Cost[i, j]$ 表示弧 $\langle v_i, v_j \rangle$ 的权值, 如果从 v_i 到 v_j 不连通, 则 $Cost[i, j] = \infty$;



- 2) 引进一个辅助向量 $Dist$, 每个分量 $Dist[i]$ 表示从起始点 v_{i0} 到每个终点 v_i 的最短路径长度, 并设定该向量的初始值为 $Dist[i] = Cost[i0, i]$, $v_i \in V$, 令 S 为已经找到的从起点出发的最短路径的终点的集合, 初值为 $\{v_{i0}\}$
- 3) 选择 V_j , 使得 $Dist[j] = \min\{Dist[i] \mid v_i \in V - S\}$, v_j 就是当前求得的一条从 v_{i0} 出发的最短路径的终点, 而 $Dist[j]$ 则是其路径长度, 令 $S = S \cup \{v_j\}$
- 4) 修改从 v_{i0} 出发到集合 $V - S$ 中任意一顶点 v_k 的最短路径长度, 如果 $Dist[j] + Cost[j, k] < Dist[k]$, 则修改 $Dist[k]$ 为: $Dist[k] = Dist[j] + Cost[j, k]$
- 5) 重复第三、四步操作共 $n-1$ 次, 由此求得从 v_{i0} 出发的到图上各个顶点的最短路径是依路径长度递增的序列。在实际应用中, 采用 Dijkstra 算法计算两点之间的最短路径和求从一点到其它所有点的最短路径所需要的时间是一样的, 算法时间复杂度为 $O(n^2)$

终点	从 v_0 到各终点的D值和最短路径求解过程				
	I=1	I=2	I=3	I=4	I=5
V_1	∞	∞	∞	∞	∞
V_2	10 (V_0, V_2)				无
V_3	∞	60 (V_0, V_2, V_3)	50 (V_0, V_4, V_3)		
V_4	30 (V_0, V_4)	30 (V_0, V_4)			
V_5	100 (V_0, V_5)	100 (V_0, V_5)	90 (V_0, V_4, V_5)	60 (V_0, V_4, V_3, V_5)	
V_j	V_2	V_4	V_3	V_5	
S	$\{V_0, V_2\}$	$\{V_0, V_2, V_4\}$	$\{V_0, V_2, V_4, V_3\}$	$\{V_0, V_2, V_3, V_4, V_5\}$	

➤ 栅格数据空间分析

1. 栅格数据处理与分析通常使用线性代数的二维数字矩阵分析法作为数据分析的数学基础。栅格数据的空间分析方法具有自动分析处理较为简单、分析处理模式化很强的特点。包括: 聚类聚合分析、叠置分析、追踪分析窗口分析等。
2. 栅格数据的聚类分析: 将栅格数据系统经某种变换而得到具有新含义的栅格数据系统

的数据处理过程。包括单一层面的栅格数据处理和多个层面的栅格数据处理。

- 1) 单一层面的栅格数据聚类分析：指根据设定的某种聚类条件对单一层面的栅格数据进行有选择的信息提取，从而建立新的栅格数据系统的方法；
- 2) 多层面的栅格数据聚类分析：实际应用中，常常需要对多层面栅格数据构成的栅格数据集进行聚类分析。每个栅格图层代表某个专题：土地利用、土壤、道路、河流或高程，或者是遥感图像的某波段的光谱值。

3. K 均值聚类算法：

设栅格数据集 $X = \{x_1, x_2, \dots, x_n\} \subset R^s$ 为 s 维的特征矢量， s 表示栅格数据的层数， n 表示每层的栅格单元数。 $x_i = (x_{i1}, x_{i2}, \dots, x_{is})$ 为栅格单元 x_i 的特征矢量或模式矢量，表示栅格单元 i 的 s 个栅格层面的属性值。

- 1) 第一步：选取聚类中心，适当地选取 k 个类的初始中心 $Z_1^{(1)}, Z_2^{(1)}, \dots, Z_k^{(1)}$ 。
- 2) 第二步：计算栅格单元到聚类中心的距离，在第 m 次迭代中，对任一栅格单元 X ，计算其到每个聚类中心的距离。距离计算采用欧式距离法，距离计算公式：

$$D_{ij} = \|X_i - Z_j^{(l)}\| = \sqrt{\sum_{p=1}^s (x_{ip} - z_{jp})^2}$$

对于所有的 $i \neq j, i = 1, 2, \dots, k$ ，如果 $\|X - Z_j^{(m)}\| < \|X - Z_i^{(m)}\|$ ，则 $X \in S_j^{(m)}$ ，

其中 $S_j^{(m)}$ 是以 $Z_j^{(m)}$ 为中心的类。

- 3) 第三步：由第二步结果计算 $S_j^{(m)}$ 新的聚类中心 $Z_j^{(m+1)}$

$$Z_j^{(m+1)} = \frac{1}{N_j} \sum_{X \in S_j^{(m)}} X$$

N_j 为 $S_j^{(m)}$ 类中的样本数； $Z_j^{(m+1)}$ 是按照使 J 最小的原则(最小平方误差准则)确定的。 J 的表达式为：

$$J = \sum_{j=1}^k \sum_{X \in S_j^m} \|X - Z_j^{(m)}\|^2$$

- 4) 第四步：迭代条件：对于所有的 $i = 1, 2, \dots, k$ ，如果 $Z_j(m+1) = Z_j(m)$ ，或者二者的差值小于一个很小的阈值，则迭代结束，否则跳转到第二步继续迭代。



4. **栅格数据的聚合分析**：指根据空间分辨率和分类表，进行数据类型的合并或转换以实现空间地域的兼并。**【将聚类得到的类合并】**将较复杂的类别转换为较简单的类别，并且常以较小比例尺的图形输出。当从小区域到大区域的制图综合变换时常需要使用这种分析处理方法。
5. 栅格数据的叠置分析可分为：简单的视觉信息复合和较复杂的叠加分类模型两种
6. **视觉信息复合**：是将不同专题的内容叠加显示在结果图件上，以便系统使用者判断不同专题地理实体的相互空间关系，获得更为丰富的信息。

地理信息系统中视觉信息复合：

- 1) 面状图、线状图和点状图之间的复合；
- 2) 面状图区域边界之间或一个面状图与其它专题区域边界之间的复合；
- 3) 遥感影像与专题地图的复合；
- 4) 专题地图与数字高程模型复合显示立体专题图；
- 5) 遥感影像与 DEM 复合生成真实三维地物景观。

视觉信息的叠加不产生新的数据层面，只是将多层信息复合显示，便于分析。

7. 叠加分类模型根据参加复合的数据平面各类别的空间关系重新划分空间区域，使每个空间区域内各空间点的属性组合一致。
 - 1) 叠加结果**生成新的数据平面**，该平面图形数据记录了重新划分的区域
 - 2) 属性数据库结构中**包括了原来的参加复合的数据平面**的属性数据库中所有的数据项
 - 3) 逻辑判断：与、或、异或、非。
 - 4) 数学运算：①算数运算：图层相加减；②函数运算：图层依照某种函数关系组合
8. 栅格数据的追踪分析：对于特定的栅格数据系统，由某一个或多个起点，按照一定的追踪线索进行目标追踪或轨迹追踪，以便进行信息提取的空间分析方法。

➤ **三维数据空间分析**

1. **空间趋势**反映的是空间物体在空间区域上变化的主体特征，它忽略局部变异以揭示总体规律。
2. **趋势面分析**：
 - 1) **趋势面**：是揭示面状区域上连续分布现象空间变化规律的理想工具，也是实际当

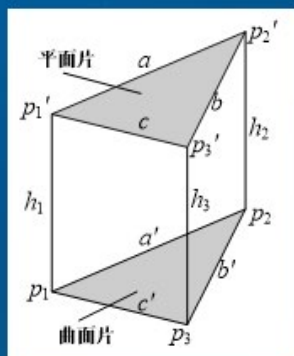
中经常使用的描述空间趋势的主要方法。

- 2) **趋势面分析**根据空间的抽样数据，拟合一个数学曲面，用该数学曲面来反映空间分布的变化情况。在数学上，趋势面分析问题实际上就是曲面拟合问题

3. 表面积计算：

- 1) 分块曲面拟合：曲面表面积由分块曲面表面积之和给出。
- 2) 全局拟合的曲面：将计算区域分成若干规则单元，先计算每个单元的面积，再累积计算总面积。
- 3) 三角形格网表面积：利用海伦公式计算面积：

三角形格网上的表面积计算



利用海伦公式计算面积：

$$\begin{cases} S = [P(P-a)(P-b)(P-c)]^{1/2} \\ P = (a+b+c) / 2 \end{cases}$$

a, b, c 的长度必须根据数据点 P_1, P_2, P_3 上的数据值 h_1, h_2, h_3 以及 $\triangle P_1P_2P_3$ 的边长 a', b', c' 计算：

$$\begin{cases} a = (a'^2 + (h_1 - h_2)^2)^{1/2} \\ b = (b'^2 + (h_2 - h_3)^2)^{1/2} \\ c = (c'^2 + (h_3 - h_1)^2)^{1/2} \end{cases}$$

- 4) 正方形格网上表面积的计算：曲面拟合重积分法和分解为三角形的方法

曲面拟合重积分方法

- ▶ 根据数学分析，某定义域 A 上的空间单值曲面 $Z=f(x, y)$ 的面积由以下重积分计算：

$$S = \iint_A (1 + f_x^2 + f_y^2)^{1/2} dx dy$$

- ▶ 该式无法直接计算，常用的方法是近似计算。比较常用的方法是 **抛物线求积方法，亦称辛卜生方法。**
- ▶ 这一方法的基本思想是先用 **二次抛物面逼近面积计算函数**，进而将抛物面的 **表面积计算转换为函数值计算。**

分解为三角形的方法

- ▶ 将正方形格网 DEM 的每个格网分解为三角形，利用计算三角形的表面积的计算公式，即海伦公式分别计算分解的三角形的面积，然后累加即得到正方形格网 DEM 的面积。

$$S = \sqrt{P(P-D_1)(P-D_2)(P-D_3)}$$

$$P = \frac{1}{2}(D_1 + D_2 + D_3)$$

$$D_i = \sqrt{\Delta X^2 + \Delta Y^2 + \Delta Z^2} \quad (1 \leq i \leq 3)$$

D_i : 第 i 对三角形两顶点之间的表面距离；

S : 三角形的表面积；


P : 三角形周长的一半。

4. 体积计算:

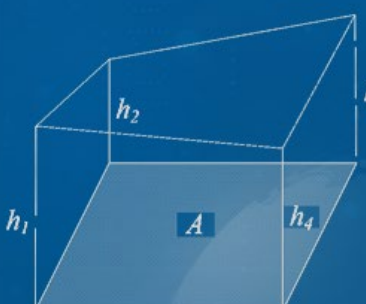
1) 基本思想: 以基底面积(三角形或正方形)乘以格网点曲面高度的均值, 区域总体积是这些基本格网体积之和。


2) 定义: 指空间曲面与某基准平面之间的空间的体积。

【基准平面是一水平面; 基准平面的高度不同, 当高度上升时, 空间曲面的高度可能低于基准平面, 出现负的体积; 在对地形数据的处理中, 体积为正示, 工程中称之为“挖方”, 体积为负时称之为“填方”。】

 **基于正方形格网的体积计算**


【体积计算公式】

$$V = S_A(h_1 + h_2 + h_3 + h_4) / 4$$


 **基于三角形格网的体积计算**

【体积计算公式】

S_A 是基底格网三角形A的面积,
三角形格网的基本格网的体积计算公式:

$$V = S_A(h_1 + h_2 + h_3) / 3$$


5. 坡度和坡向计算:

1) 坡度: 某点在曲线上的法线方向与垂直方向的夹角, 是地面特定点高度变化比率的度量。【反映斜坡的倾斜程度, 是一个具有方向与大小的矢量】

2) 坡度的表示方法: 有百分比法、度数法、密位法和分数法四种, 其中以百分比法和度数法较为常用。

a) 百分比法: 最为常用的方法, 即两点的高程差与其水平距离的百分比。

计算公式: 坡度 = (高程差/水平距离)×100%, $i = h/l \times 100\%$ 。

b) 度数法: 用度数来表示坡度, 利用反三角函数计算而得。

计算公式:

$\tan\alpha(\text{坡度}) = \text{高程差}/\text{水平距离}$, $\alpha(\text{坡度}) = \arctan(\text{高程差}/\text{水平距离})$

- 3) 坡向：法线的正方向在平面上的投影与正北方向的夹角，即法方向水平投影向量的方位角。取值范围从零方向(正北方向)顺时针到 360 度。【反映斜坡所面对的方向。】

- 4) 基于规则格网 DEM 的坡度坡向计算：由单元标准矢量的倾斜方向和倾斜量，计算每个单元的坡度和坡向。

标准矢量：指垂直于格网单元的有向直线，设标准矢量为(n_x, n_y, n_z)。

该格网单元的坡度 S 为：

$$S = \frac{\sqrt{n_x^2 + n_y^2}}{n_z}$$

格网单元的坡向 D 为

$$D = \arctan(n_x/n_y)$$

在实际计算时，通常是用 3×3 的移动窗口来计算中心单元的坡度和坡向。

计算时考虑邻接单元的影响有不同方式：

- a) Ritter 算法：只考虑直接与中心点单元相邻的 4 个单元。

中心点 e 的坡度：

$$S_e = \sqrt{(e_1 - e_3)^2 + (e_4 - e_2)^2} / 2d$$

中心点 e 的坡向：

$$D_e = \arctan[(e_4 - e_2)/(e_1 - e_3)] + 90^\circ$$

e_i：相邻单元值；d：单元大小；

(e₁-e₃)：x 方向的高差；(e₄-e₂)：y 方向的高差

- b) Horn 算法：考虑与中心单元相邻的 8 个相邻单元，直接邻接单元(e₂, e₄, e₆, e₈)的权值为 2，其它 4 个单元 (e₁, e₃, e₅, e₇)的权值为 1。

中心点 e 的坡度：

$$S_e = \frac{\sqrt{[(e_1 + 2e_4 + e_6) - (e_3 + 2e_5 + e_8)]^2 + [(e_6 + 2e_7 + e_8) - (e_1 + 2e_2 + e_3)]^2}}{8d}$$

中心单元 e 的坡向：

$$D_e = \arctan \frac{(e_6 + 2e_7 + e_8) - (e_1 + 2e_2 + e_3)}{(e_1 + 2e_4 + e_6) - (e_3 + 2e_5 + e_8)}$$

- 5) 基于不规则三角网(TIN)的坡度坡向计算

设三角面的三个节点坐标分别为 $E_1(x_1, y_1, z_1)$, $E_2(x_2, y_2, z_2)$ 和 $E_3(x_3, y_3, z_3)$

使用双向标准矢量，该矢量垂直于三角面，标准矢量为 $\vec{E_1E_2} = (x_2 - x_1, y_2 -$

$y_1, z_2 - z_1$ 和 $\vec{E_1 E_3} = (x_3 - x_1, y_3 - y_1, z_3 - z_1)$ 的向量积。

标准向量的三个分量为：

$$n_x: (y_2 - y_1)(z_3 - z_1) - (y_3 - y_1)(z_2 - z_1)$$

$$n_y: (z_2 - z_1)(x_3 - x_1) - (z_3 - z_1)(x_2 - x_1)$$

$$n_z: (x_2 - x_1)(y_3 - y_1) - (x_3 - x_1)(y_2 - y_1)$$

带入以下两式可计算坡度和坡向：

$$S = \sqrt{n_x^2 + n_y^2} / n_z$$

$$D = \arctan(n_x / n_y)$$

6) 基于等高线的坡度坡向计算

6. **剖面分析：**以数字高程模型（DEM）为基础构造某一个方向的剖面，以线代面，概括研究区域的地势、地质和水文特征；

地形剖面线是剖面分析的基础。地形剖面线根据所选剖面与数字地形图上地形表面的交点反应地形的起伏情况。

地形剖面线的生成方法：

1) 基于规则格网的剖面线生成算法：

- 确定剖面线的起止点：由坐标确定，或用鼠标在三维场景中选择决定；
- 计算剖面线与经过网格的所有交点，内插出各交点的坐标和高程，将交点按离起始点的距离排序；
- 顺序连接相邻交点，得到剖面线；
- 选择一定的垂直比例尺和水平比例尺，以离起始点的距离为横坐标，以各点的高程值为纵坐标绘制剖面图。

2) 基于不规则三角网的剖面线生成方法：

- 用剖面所在的直线与 TIN 中的三角面的交点得到剖面线；
- 先利用 TIN 中各三角形构建的拓扑关系快速找到与剖面线相交的三角面，再进行交点的计算，这样可以提高运算速度；
- 最后以距离起始点的距离为横坐标，以各点的高程值为纵坐标绘制剖面图。

7. **可视性分析：**也称为视线图分析、或通视分析，用于描述通视情况。实质上属于对地形进行最优化处理的范畴。

1) 两点之间的可视性(inter-visibility)分析

- a) 基于格网 DEM 的通视分析：将格网点作为计算单位，把点对点的通视问题简化为 DEM 格网与某一地形剖面线(视线)的相交问题。
- 2) 可视域(viewshed)分析
- 3) 可视性分析的基本用途：
 - a) 可视查询：指对于给定的地形环境中的目标对象(或区域)，确定从某个观察点观察，该目标对象是全部可视还是某一部分可视。
 - b) 表面路径问题：一个与可视域和水平可视计算都相关的应用问题。解决地形环境中与通视相关的路径设置问题。

➤ 空间数据统计分析

1. 多元统计分析：研究客观事物中多个变量（或多个因素）之间相互依赖的统计规律性的空间数据分析方法。

1) 主成分分析法(PCA, Principal Components Analysis)

- a) 是把原来多个变量划为少数几个综合指标的一种统计分析方法。
- b) 特点：将众多要素的信息压缩表达为若干具有代表性的合成变量，可以克服变量选择时的冗余和相关，选择信息最丰富的少数因子进行各种分析，构造应用模型。
- c) 思路：设有 n 个样本、 p 个变量，将原始数据转换成一组新特征值“主成分”，主成分是原变量的线性组合且具有正交特征，将 x_1, x_2, \dots, x_p 综合成 $m(m < p)$ 个指标 z_1, z_2, \dots, z_m 。综合指标 z_1, z_2, \dots, z_m 分别称作原指标的第一，第二， \dots ，第 m 主成分； z_1 在总方差中占比最大， z_2, z_3, \dots, z_m 的方差依次递减

2) 层次分析法(Antalytic Hierarchy Process, AHP)

- a) 把思维过程层次化、数量化，是一种定性和定量分析相结合的方法。
- b) 特点：
 - i. 层次分析法把相互关联的要素按隶属关系分为若干层次，请有经验的专家对各层次各因素的相对重要性给出定量指标，
 - ii. 利用数学方法综合专家意见给出各层次各要素的相对重要性权值，作为

综合分析的基础。判断矩阵中的值 a_{ij} 可通过专家打分建立。

3) 系统聚类分析法:

- a) 根据多种地学要素对地理实体进行类别划分的分析方法。
- b) 准则: 类间差异最大、类内差异最小; 根据实体间的相似程度, 逐步合并若干类别。
- c) 根据差异性: 欧氏距离、马氏距离、切比雪夫距离、兰氏距离、绝对距离等
- d) 根据相似性: 皮尔逊相关系数(Pearson Correlation coefficient)、斯皮尔曼秩相关系数 (Spearman Rank Correlation)、肯德尔秩相关系数 (Kendall Rank Correlation)、余弦相似度(Cosine Similarity)

4) 判别分析法:

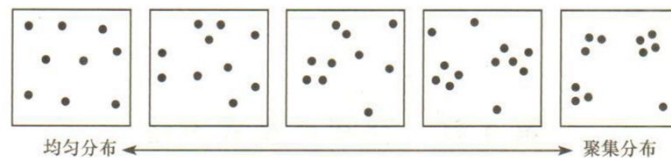
- a) 根据某一研究对象的各种特征值判别其类别归属的一种多元统计分析方法。
- b) 判别函数: 根据已知的地理特征值进行线性组合, 构成一个线性判别函数

$$Y = c_1 * x_1 + c_2 * x_2 + \dots + c_k * x_k + \dots + c_m * x_m, c_k (k=1, 2, \dots, m) \text{ 为}$$

判别系数, 反映各要素或特征值作用方向、分辨能力和贡献率的大小; 确定了 c_k , 判别函数 Y 就确定了

2. 空间点模式分析

点模式的三种基本类型: 聚集分布、随机分布、均匀分布



3. 基于密度的方法

1) 基于密度的方法——样方分析法(quadrat analysis, QA):

- a) 研究空间点模式的最常用的直观方式。
- b) 基本思想: 通过空间上点分布密度的变化探索空间分布模式。
- c) 一般过程:
 - i. 将研究区域划分为规则的正方形网格区域;
 - ii. 统计落入每个网格中点的数量;
 - iii. 统计出包含不同数量点的网格数量的频率分布;
 - iv. 将观测得到的频率分布和已知的频率分布或理论上的随机分布(如泊松分

布)作比较, 判断点模式的类型。

2) 基于密度的方法——核密度估计法(kernel density estimation, KDE)

- a) 一种使用事件的空间密度分析表示空间点模式的方法。
- b) 基本思想: 认为地理事件可以发生在空间的任何位置上, 但是不同位置上事件发生的概率不一样。点密集的区域事件发生的概率高, 点稀疏的区域事件发生的概率低。
- c) 空间点模式在点 S 上的密度或强度是可测度的, 一般通过测量定义在研究区域中单位面积上的事件数量来估计。最简单的方法是在研究区域中使用滑动的圆来统计出落在圆域内的事件数量, 再除以圆的面积, 得到估计点 S 处的事件密度。

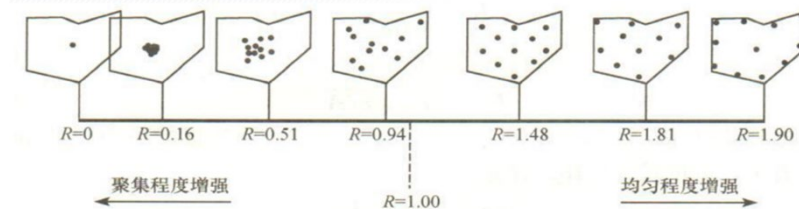
4. 基于距离的方法

1) 基于距离的方法——最邻近指数法 (NNI: Nearest Neighbor Index)

- a) 使用最邻近的点对之间的距离描述分布模式, 形式上相当于密度的倒数 (每个点代表的面积), 表示点间距。
- b) 最邻近距离: 指任意一个点到其邻近的点对之间的距离。
【☆点对之间的最邻近距离不是相互的, 即: j 点是第 i 个点的最邻近点, 但 i 不一定是 j 的最邻近点。在完全空间随机模式 (CSR) 模式中, 超过 60% 的最邻近是相互的邻近。】
- c) 首先计算观测模式和已知模式的最邻近点对之间的平均距离 \bar{d}_{min} , 然后根据观测模式的 NNI 与 CSR 模式的 NNI 比较进行判断, 确定分布模式。
 - i. 如果观测模式的最邻近距离 **大于随机分布模式的最邻近距离**, 则观测模式 **趋向于均匀分布**;
 - ii. 如果观测模式的最邻近距离 **小于随机分布模式的最邻近距离**, 则观测模式 **趋向于聚集分布**。
 - iii. 具体判断方法: 在 CSR 模式中, 同样计算平均的最邻近距离, 将其作为期望 $E(d_{min})$; 计算最邻近指数 $R: R = \frac{\bar{d}_{min}}{E(d_{min})} = 2\bar{d}_{min}\sqrt{n/A}$ 。
 - ① $R=1$, 说明观测事件过程来自于完全空间随机模式 CSR, 属于随机分布
 - ② $R<1$, 说明不是完全空间随机模式 CSR, 表明大量事件点在空间上相

互 接近，属于空间聚集模式

③ $R > 1$ ，说明不是完全空间随机模式 CSR，由于点之间的最邻近距离大于 CSR 过程的最邻近距离，空间点之间是相互排斥的，趋于均匀分布模式



☆ 极端聚集：所有事件发生在研究区域的同一个位置， $R=0$

☆ 极端均匀：均质区域上邻近的 3 个点构成等边三角形，空间被正六边形划分，点位于正六边形的中心，是克里斯泰勒中心地分布模式，

$R=2.149$

5. 空间相关性统计概述

- 1) 空间相关性统计是研究面状单元空间关系作用下的变量值的空间模式。
- 2) 面积单元之间的邻接与否、距离远近等对于变量的空间分布具有重要影响。
- 3) **空间接近性**：面积单元之间的“距离关系”，基于距离的空间接近性测度就是使用面积单元之间的距离定义邻接性。
- 4) **测度面积单元距离的方法——边界邻接法**：面积单元之间具有共享的边界，被称为是空间邻接的。直接邻接、间接邻接、多重邻接。
- 5) **测度面积单元距离的方法——重心距离法**：面积单元的重心或中心之间的距离小于某个指定的距离，则面积单元在空间上是邻接的。
- 6) **空间权重矩阵**：空间邻接性的定量化测度。
 - a) 假设研究区域中有 n 个多边形，任何两个多边形都存在一个空间关系，这样就有 $n \times n$ 对关系。需要 $n \times n$ 的矩阵存储这 n 个面积单元之间的空间关系。
 - b) 空间权重矩阵的类型：
 - i. 左右相邻权重： $w_{ij} = \begin{cases} 1, & \text{区域 } i \text{ 和 } j \text{ 的邻接为左右邻接} \\ 0, & \text{其他} \end{cases}$
 - ii. 上下相邻权重： $w_{ij} = \begin{cases} 1, & \text{区域 } i \text{ 和 } j \text{ 的邻接为上下邻接} \\ 0, & \text{其他} \end{cases}$



iii. Queen 权重: $w_{ij} = \begin{cases} 1, & \text{区域 } i \text{ 和 } j \text{ 有公共边或公共点} \\ 0, & \text{其他} \end{cases}$

iv. 基于距离的权重: $w_{ij} = \begin{cases} 1, & \text{区域 } i \text{ 和 } j \text{ 的距离小于 } d \\ 0, & \text{其他} \end{cases}$

v. Dacey 权重: $w_{ij} = d_{ij} \times \alpha_i \times \beta_{ij}$

d_{ij} 对应二进制连接矩阵元素, 取值为 1 或 0;

α_i 是单元 i 的面积占整个空间系统的所有单元的总面积的比例;

β_{ij} 为单元 i 与单元 j 共享的边界长度占 i 单元总边界长度的比例。

vi. 阈值权重: $w_{ij} = \begin{cases} 0, & i = j \\ a1, & d_{ij} < d \\ a2, & d_{ij} \geq d \end{cases}$

7) **空间自相关**: 描述空间域中某位置上的变量与其**邻近位置上同一变量**的相关性。

是空间数据的重要性质。

a) 如果邻接位置上相互间数值接近, 空间模式表现出**正空间自相关**;

b) 如果邻接位置上相互间的数值不接近, 空间模式表现出**负空间自相关**。

c) 空间自相关性的指标类型:

i. 全局指标: 用于探测整个研究区域的空间模式, 用一个值来反映该区域的自相关程度;

ii. 局部指标: 计算每一个空间单元与邻近单元就某一属性的相关程度

6. **全局空间自相关分析**: 用于探测整个研究区域的空间模式, 用一个值来反映该区域的自相关程度。

1) **全局空间关联指标——全局 Moran's I 统计量**

全局 Moran's I 统计量: 衡量相邻的空间分布对象及其属性取值之间的关系

$$I = \frac{N \cdot \sum_i \sum_j w_{ij} \cdot (y_i - \bar{y})(y_j - \bar{y})}{(\sum_i \sum_j w_{ij}) \cdot \sum_i (y_i - \bar{y})^2}$$

w_{ij} : 衡量空间事务之间关系的权重矩阵, 一般为对称矩阵。

【 $w_{ij}=0$: 空间自相关研究同一属性不同地理位置的相关性, 同一地点的属性相关性没有意义, 因此 $w_{ij}=0$ 。】

Moran's I 的取值范围: $[-1, 1]$

①正值: 该空间事物的属性取值具有正相关性

- ②负值: 该空间事物的属性取值具有负相关性;
- ③零值: 空间事物的属性取值不存在空间相关, 即空间随机分布。

2) 全局空间关联指标——全局 Geary's C 统计量

与全局 Moran's I 相似, 区别在于: 分子的交叉乘积项不同, 即测量邻近空间位置观察值近似程度的方法不同。全局 Moran's I 的交叉乘积项比较的是邻近空间位置的观察值与均值偏差的乘积。全局 Geary's C 比较的是邻近空间位置的观察值之差。

全局 Geary's C 统计量:

$$C = \frac{\sum_i^N \sum_j^N w_{ij} \cdot (y_i - y_j)(y_i - y_j)}{2(\sum_i^N \sum_j^N w_{ij}) \cdot \sum_i^N (y_i - \bar{y})^2 / (N - 1)}$$

全局 Geary's C 统计量的取值范围为[0, 2]:

- ①当 $0 < C < 1$ 时, 表示具有该属性取值的空间事务分布具有正相关性;
- ②当 $1 < C < 2$ 时, 表示具有该属性取值的空间事务分布具有负相关性;
- ③当 $C \approx 1$ 时, 表示不存在空间相关性。

3) 全局空间关联指标——全局 Getis-Ord G 统计量

首先设定一个距离阈值, 在给定阈值的情况下, 决定各数据的空间关系, 然后分析其属性乘积来衡量这些空间对象取值的空间关系。

Getis-Ord G 统计量直接采用邻近空间位置的观察值之积来测量其近似程度

全局 Getis-Ord G 统计量:

$$G(d) = \frac{\sum_i^N \sum_{j,j \neq i}^N w_{ij} \cdot (d) y_i y_j}{\sum_i^N \sum_{j,j \neq i}^N y_i y_j}$$

y_i 为各数据的属性值, $w_{ij}(d)$ 为给定距离阈值 d 下 i, j 两者空间关系的权重矩阵。

Getis's G 的统计空间自相关性通过得分检验进行:

$$Z(G) = \frac{(G(d) - E(G(d)))}{\sqrt{\text{var}(G)}}$$

- ①当 $Z(G)$ 为正值时, 表示属性取值较高的空间对象存在空间聚集关系;
- ②当 $Z(G)$ 为负值时, 表示属性取值较低的空间对象存在空间聚集关系;

全局 Getis-Ord G 统计量的两个区：

- a) **热点区 (hot spots):** 当观察值大的空间位置相互邻近时, 这种正空间自相关通常称为热点区;
- b) **冷点区 (cold spots):** 当观察值低的空间位置相互邻近时, 这种正空间自相关通常称为冷点区。
- c) ☆全局 Getis-Ord G 的优势: 可以非常好地区分这两种不同的正空间自相关。

7. 局部空间自相关分析: 计算每一个空间单元与邻近单元就某一属性的相关程度。

1) 局部 Moran's I 统计量:

$$I_i = \frac{y_i - \bar{y}}{S^2} \sum_{j=1}^N w_{ij} (y_j - \bar{y}), U(I_i) = \frac{I_i - E(I_i)}{\sqrt{\text{var}(I_i)}}; S^2 = \sum_{j=1, j \neq i}^N y_j^2 / (N - 1) - \bar{y}^2$$

I_i 为第 i 个分布对象的全局相关性系数;

$E(I_i)$ 表示空间位置 i 的观测值的数学期望;

$\text{var}(I_i)$ 表示空间位置 i 的观测值的方差。

- a) 局部 Moran's I 的值大于数学期望, 并且有统计学意义时, 提示存在局部的正空间自相关;
- b) 局部 Moran's I 的值小于数学期望, 提示存在局部的负空间自相关。

2) 局部 Geary's C 统计量:

$$C_i = \sum_j w_{ij} \left(\frac{x_i - \bar{x}}{\sigma} - \frac{x_j - \bar{x}}{\sigma} \right)^2, U(C_i) = \frac{C_i - E(C_i)}{\sqrt{\text{var}(C_i)}}$$

参数意义与 Moran's I 相同

- a) 局部 Geary's C 的值小于数学期望, 并且有统计学意义时, 提示存在局部的正空间自相关;
- b) 局部 Geary's C 的值大于数学期望, 提示存在局部的负空间自相关。

3) 局部 Getis-Ord G 统计量:

同全局 Getis-Ord G 一样, 只能采用距离定义的空间邻近方法生成权重矩阵

$$C_i(d) = \frac{\sum_j w_{ij}(d)x_j}{\sum_j x_j}, U(G_i) = \frac{G_i - E(G_i)}{\sqrt{\text{var}(G_i)}}$$

- a) 当局部 Getis-Ord G 的值大于数学期望, 并且有统计学意义时, 提示存在“热点区”;
- b) 当局部 Getis-Ord G 的值小于数学期望, 提示存在“冷点区”;

4) 三种指标的比较

- a) 局部 Moran's I 和 局部 Geary's C: 可以很好地识别正空间自相关、负空间自相关; 但是不能区分“热点”和“冷点”这两种不同的正空间自相关。
- b) 局部 Getis-Ord G: 可以很好地区分“热点”和“冷点”这两种不同的正空间自相关; 但是识别负空间自相关时效果较差。

8. 地统计分析

1) 地统计学(Geostatistics): 也称为地质统计学, 是一门以区域化变量理论为基础,

以变异函数为主要工具, 研究那些分布于空间上既有随机性又有结构性的自然或社会现象的科学。

2) 主要包括三个主要内容: 区域化变量理论、变异函数模型、克里金估计方法

3) 地统计分析的两个特点:

- a) 样本点的空间相关性。传统统计中不同样本点仅具有随机性, 样本点之间保持空间独立性。然而, 地统计中样本点不仅具有随机性, 同时样本点之间具有空间相关性。
- b) 一次性样本采集。传统统计分析同一空间位置处可以多次采样数据。实际地统计分析中, 样本区域中每一个空间位置多为一次采样数据。
- c) 这两个特点导致了地统计中描述空间相关性(空间结构)的变异函数和克服一次采样局限的平稳性假设的提出。

4) 区域化变量理论:

- a) 区域化变量(regionalized variable): 设 $Z(x)$ 为一随机变量, 表示在空间位置 x 处专题变量取值是随机的, 区域化变量是区域化随机变量的简称。

- b) **随机场**: $Z(X)=\{Z(x), x \in X\}$ 表示区域 X 中所有空间位置 x 处随机变量 $Z(x)$ 的集合, 又称为随机场, 随机场也可看作若干空间样本的集合。
- c) **区域化变量理论**: 揭示区域化变量 **空间结构和统计性质** 的理论, 简称为区域化变量理论, 构成了地统计学的基础。重点研究区域化随机变量的各种空间结构和统计性质。

5) **空间变异函数**:

- a) 是描述区域化随机变量空间结构的有效数学工具。
- b) **二阶平稳性假设**: 如果区域化变量 $Z(x)$ 满足下列两个条件, 则称其满足二阶平稳性假设。
- i. 在研究范围内, 区域化变量 $Z(x)$ 的期望存在且为常数, 即 $E[Z(x)] = m$
- ii. 在研究范围内, 区域化变量 $Z(x)$ 的协方差函数存在且为空间滞后 h 的函数, 与空间位置 x 无关, 即:

$$Cov[Z(x), Z(x+h)] = E\{[Z(x+h) - m][Z(x) - m]\} = E[Z(x+h)Z(x)] - m^2 = C(h)$$

【当 $h=0$ 时, 条件(2)说明了方差函数存在且为常数,

$$Var[Z(x)] = Cov[Z(x), Z(x)] = E[Z(x) - m]^2 = C(0)】$$

- c) **内蕴性假设**: 如果区域化变量 $Z(x)$ 满足下列两个条件, 则称其满足内蕴性假设。
- i. 在研究范围内, 区域化变量 $Z(x)$ 增量的期望为零, 即

$$E[Z(x+h) - Z(x)] = 0$$

- ii. 在研究范围内, 区域化变量 $Z(x)$ 增量的方差存在且为空间滞后 h 的函数, 与空间位置 x 无关, 即

$$\begin{aligned} Var[Z(x+h) - Z(x)] &= E\{[Z(x+h) - Z(x)] - E[Z(x+h) - Z(x)]\}^2 \\ &= E[Z(x+h) - Z(x)]^2 = 2\gamma(h) \end{aligned}$$

【 $\gamma(h)$ 表示区域化变量的 **变异函数或半方差函数**, 也将 $r(h)$ 称为半变异函数或半变差函数。】

- d) **变异函数的数学表示**: 变异函数是区域化变量空间结构的一种形式化表达, 是从 **增量的方差** 角度定义的, 是 **空间位置 x 和空间滞后 h** 的函数。在二阶平稳性或内蕴平稳性假设 (期望不变, 协方差或变异函数仅空间滞后相关) 下, **原始变异函数 $r_x(h)$ 规约为单纯的空间滞后 h 的 $r(h)$** , 与空间位置 x 无

关。数学表示为两个随机变量 $Z(x)$ 和 $Z(x+h)$ 之间增量的方差的一半：

$$\gamma_x(h) = \frac{1}{2} \text{Var}[Z(x+h) - Z(x)]$$

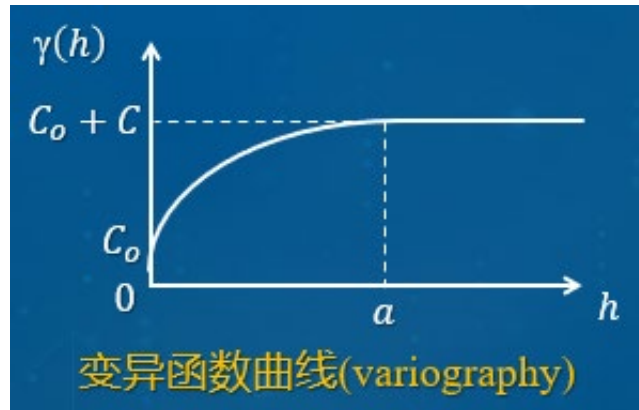
e) **变异函数曲线(variography)**: 变异函数值随着空间滞后 h 的增大而单调增加

变异函数 $\gamma(h)$ 具有三个参数 $\{a, C_0, C_0 + C\}$

块金值 C_0 : 是空间滞后为0时的变异函数值, 是测量误差和低于采样间距的随机变异的综合反映

基台值 $C_0 + C$: 当空间滞后 h 超过变程 a 时, 变异函数 $\gamma(h)$ 在一个极限值 $\gamma(\infty)$ 附近摆动, 这个极限值称为基台值 $C_0 + C$

变程 a : a 称为变程, 是变异函数达到基台值 $C_0 + C$ 时的空间滞后 h , 反映了空间自相关的最大距离



区域化变量的取值 z 由大尺度趋势 μ 、微尺度空间相关变异 r 和纯随机变异 ε 三部分构成, 即 $Z = \mu + r + \varepsilon$

μ : 期望(或平均值) μ 是一种趋势表示

r : 微尺度空间相关变异 r 为去除趋势后具有内在空间自相关性的残余值

ε : 纯随机变异 ε 为不存在空间自相关性的独立噪声(如测量误差)

测量误差和采样间距(采样尺度)以下的微尺度空间相关残余值一起构成块金值 C_0 。采样间距(采样尺度)以上的微尺度空间相关残余值的变异函数值为 C 。

6) **克里金估计方法**:

a) **克里金估计**: 利用区域化变量结构性质进行估值应用。克里金插值和克里金预测统称为克里金估计。

- b) 不同类型的克里金估计方法:
- 区域化变量满足二阶平稳性(或内蕴性)假设的**普通克里金估计**和**简单克里金估计**;
 - 区域化变量非平稳(存在漂移)的**泛克里金估计**;
 - 多个变量的**协同克里金估计**;
 - 变量服从对数正态分布的**对数克里金估计**;
 - 适用于非连续取值(包括名义数据)的**指示克里金估计**、**析取克里金估计**和**概率克里金估计**等;
 - 可以综合多个角度,全面利用区域化变量的结构性质,对单个特性建模的克里金估计进行组合,形成**普通协同克里金估计**、**协同泛克里金估计**和**协同指示克里金估计**等方法。

第八章·空间数据可视化与地图制图

➤ 地学可视化

1. 科学计算可视化

- 定义:运用计算机图形学和图像处理技术,将科学计算过程中产生的数据及计算结果转换为图形和图像显示出来,并进行交互处理的理论、方法和技术。
- 主要功能:利用复杂的多维数据生成图形或图像等,以帮助人们理解、分析以计算机方式存储的图形、图像数据等。

2. 地理空间数据可视化

- 定义:指运用地图学、计算机图形学和图像处理技术,将地学信息输入、处理、查询、分析以及预测的数据及结果采用图形符号、图形、图像,结合图表、文字、表格、视频、音频等可视化形式显示并进行交互处理的理论、方法和技术。

【**地理空间数据可视化是科学计算可视化在地学领域的特定发展。**】

3. 地学可视化的作用:

- 可视化可用于表达地理空间信息;
- 可视化能用于地理空间分析;
- 可视化可以用于数据的仿真模拟。



4. 可视化与相关学科的关系：

- 1) GIS 起源于或者说“脱胎于”计算机地图制图。GIS 的重点在于空间分析。
- 2) 计算机地图制图重点在于如何完成空间信息的可视化，即地图输出问题。
- 3) 地图是 GIS 一种主要的、重要的可视化和输出手段，但不是唯一的手段。
- 4) 随着计算机、虚拟现实技术等学科的发展，多媒体、虚拟现实技术也逐渐成为 GIS 可视化和输出方式。

➤ 地图制图

1. **地图**：按照严格的**数学法则**、利用特定的**符号系统**将地球或其他星球的空间对象，以**二维或多维、静态或动态**可视化形式**综合概括、模型模拟**等手段缩小表示在一定载体上，科学地分析认知与交流传输对象的**时空分布、质量特征及相互关系**等多方面信息的一种**图形与图像**。

- 1) 基本特征：**可量测性；直观性；易览性**
- 2) 空间数据地图可视化要素：地图符号；地图色彩；地图注记；数学要素；地理要素

2. 地图的分类：

- 1) 普通地图：反映**地表基本要素的一般特征**的地图，**以同等详细程度全面表示**地面基本自然和人文社会经济现象一般特征的地图，能比较完整反映出制图区域的基础地理特征。
 - a) 基本要素：水系、地形、地貌、土质、植被、居民地、交通网、境界线以及独立地物等
 - b) 特点：标准化，国家相关部门对基本比例尺的地形图的规格、大小、内容及图式都作了严格的规定。
- 2) 专题地图：根据实际任务需要，**着重反映**自然或社会现象中的**某一种或几种专业要素**的地图，其重点表现某种主题内容。

特点：

- a) **不属于**国家基本比例尺地图约定范畴；
- b) 针对不同的服务对象，其对应的**地图主题内容不同**；
- c) 用户可以根据项目的需要**自行设计和决定**专题地图的比例尺、投影坐标、图幅、非空间属性值的**符号化表示方法**等。



3. **地图基本要素**：标题；专题内容（地理要素）；地图符号；地图注记；坐标网；控制点；比例尺；定向；地图颜色；图例；地图参考资料

- 1) 地图符号：地理要素特征的抽象，是地图语言的主体内容。

分类：点状符号、线状符号、面状符号

- 2) 地图注记：名称注记、说明注记、数字注记、图外整饰注记

- 3) 坐标网：经纬线网、方里网（每隔整公里绘的坐标纵线和坐标横线）

- 4) 比例尺：地图的缩小程度，是地图上某一线段长度与实地距离的比值

根据地图投影变形情况，地图比例尺分为：

- a) 主比例尺：在投影面上没有变形处的比例尺

- b) 局部比例尺：在投影面上有变形处的比例尺

- 5) 地图定向：地图的三北方向，一般通过坐标网的方向来体现

- a) 真北方向：过地面上任意一点，指向北极的方向

【一般地图都采用真北方向定向】

- b) 坐标北方向：坐标纵坐标递增的方向

- c) 磁北方向：磁北针所指的方向

- 6) 地图颜色：应用色相、亮度和饱和度的不同变化和组合，结合人们对色彩感受的心理特征，建立起色彩与制图对象间的联系。

- a) 色相：主要表示事物的质量特征。如用蓝色表示淡水，紫色表示咸水

- b) 亮度和饱和度：表示事物的数量特征和重要程度。用较浓艳的颜色表示重要地物，用浅淡颜色表示次要事物。

4. 地图符号——点状符号

- 1) 定义：不依比例表示的小面积地物或点状地物符号

- 2) 特点：

- a) 点符号的图形固定，不随它在图幅中的位置的变化而变化

- b) 点都有确定的定位点和方向性

- c) 图形大都比较规则，由几何图形构成，简单、美观、形象，易用数学公式表示

5. 地图符号——线状符号

- 1) 定义：长度在图上依比例尺表示而宽度不用比例尺表示的符号

2) 特点:

- a) 都有一条有形或无形的定位线
- b) 可以进一步划分为曲线、直线、虚线、平行线、沿定位线连续配置点符号等
- c) 可以进一步分解成具有单一特征的线状符号, 即一线状符号可以由若干条具有单一特征的线状符号组成

6. 地图符号——面状符号

1) 定义: 在二维图上各方向都能依比例尺表示的符号

2) 特点:

- a) 有一条有形或无形的封闭的轮廓线
- b) 多数面状符号是在轮廓线范围内配置不同的点状符号、绘阴影线或着染颜色

7. 专题地图表示方法

1) 定点符号法:

- a) 基本思想: 使用不同形状、大小、颜色的符号表示点状分布的对象
- b) 用途: 主要用来表示有精确定位的点状地物

2) 线状符号法:

- a) 基本思想: 采用线状符号在地图上反映线状分布的地理现象
- b) 线状符号要素: 包括不同颜色和形状的线划、箭头、条带

3) 运动线法:

- a) 基本思想: 用矢量符号和不同宽度、颜色的条带表示空间现象移动的方向、路线和数量、质量特征等
- b) 用途:
 - i. 自然现象: 洋流、寒潮、鱼类回游、台风等运动途径
 - ii. 社会现象: 移民、货物运输、资本输入等方向
- c) 方法:
 - i. 用运动线的粗细反映现象分布的强度
 - ii. 长短反映重复出现的次数
 - iii. 用不同颜色或形状代表不同时期发生的现象

4) 等值线法:

- a) 基本思想: 用等值线形式表示布满整个制图区域的面状现象的专题图方法

- b) 用途：用来连续分布并逐渐变化的制图现象的数量特征
- 5) 质底法：
- a) 基本思想：底色法，在区域界线或类型范围内渲染颜色或填绘晕线、花纹，以显示布满制图区域专题要素的质量差别。
- b) 应用：编制各种类型图或区划图
- 6) 范围法：
- a) 基本思想：表示呈间断成片分布的面状对象
- b) 用途：适用于表示集中分布在一定面积上，而在其他制图地区不经常出现的自然和社会经济现象。如森林、沼泽、某种农作物的分布等

	范围法	质底法
所表达事物 是否布满制图区	否	是
符号能否重叠	可	否
表达事物侧重点	分布范围	质量特征

✧ 范围法与质底法的区别：

范围法所表达的事物不布满制图区，符号可以重叠，表达事物侧重于分布范围。

质底法所表达的事物布满制图区，符号不可以重叠，表达事物侧重于质量特征。

- 7) 点值法
- a) 基本思想：用一定大小和形状的点群来反映制图区域中分散的、复杂分布的现象
- b) 用途：在制图现象分布的范围内用大小相同的点群表示其分布特征。点数多少对应于该现象在表示范围内的发展程度，如人口、动物分布等
- 分级统计图法和图表统计图法
- a) 分级统计图法：用颜色深浅或符号疏密表示制图现象统计差异
- b) 图表统计图法：用柱状图或其他图形符号表示各区域单元的统计数据。
- c) 两种方法的差别：
- 分级统计图法：反映制图现象分布强弱，必须使用相对值指标；
- 图表统计图法：直接采用绝对值指标

- 9) 分区统计图表法
- 基本思想：在各分区单元内按统计数据描绘为不同形式的统计图表，置于相应的区划单元内，反映各区划内现象的总量、构成和变化
 - 用途：反映不同区划单元内特征属性的区别
8. **地图设计**：为地图制图制定表达方式、表达规则、选择表达内容和技术等技术过程
9. **地图输出方式**：
- 绘图仪输出：最简单的，最常用的方式
 - 计算机图形输出：根据绘图指令，编写绘图程序，直接驱动绘图笔绘图。
- 由 GIS 软件产生一种标准的图形文件，如 Windows 的元文件 WMF 文件，调用操作系统或者 Windows 提供的函数“播放”元文件，绘制地图。
- 所有程序不变，在需要绘图时，将图形屏幕显示的句柄改为**绘图设备句柄**
10. **电子地图制作**：
- 电子地图：以数字地图为基础，以多种媒体显示地图数据的可视化产品，是数字地图的一种可视化方式。
 - 电子地图系统：电子地图集作为传统地图的补充，不仅包含、存储了地图的全部内容，而且形成了一个信息系统。
11. **GIS 与电子地图的区别**：电子地图以地图数据库为基础，包含多种形式数据，数据具有较强的拓展性，信息存储、更新以及通信方式较为简便，便于携带和交流。电子地图检索十分方便，可以动态模拟和实时动态显示。同时可以缩短大型系列地图集的生产周期和更新周期降低生产成本。GIS 是一种兼容存储、管理、分析、显示与应用地理信息的计算机系统，是分析和处理海量地理数据的通用技术。GIS 以其强大的空间数据管理系统形象直观的应用界面、强大的空间分析能力等特点，能为显示地理空间上的物质和能量运动规律的研究提供方便、准确的管理和空间分析手段。在规划、军事、道路等许多方便应用。。
12. **GIS 与电子地图的联系**：电子地图提供对地图查询和信息可视化，而不能提供对地图数据的分析和对空间数据的挖掘。GIS 具有强大的空间分析功能，可以对海量的地理数据进行分析及提取有用的信息，但它需要准确的数据。因此电子地图可以 GIS 技术为基础,融合了数据库技术,其信息的可视化程度大大提升，表现形式更加丰富，实现三维显示。例如下图中的百度地图和 GIS 中融合电子地图。

➤ 地形数据可视化

1. 等值线法

- 1) 等值线：将特定属性取值相等的点连接起来。
- 2) 常用等值线：等高线
 - a) 等高距：等高线之间的垂直距离
 - b) 基准等高线：开始计算高程的等高线
 - c) 等高线的排列和分布模式，有效反映地形特点
- 3) 等值线绘制算法——基于规则格网绘制方法
 - a) 计算各条等值线和网格边交点的坐标值
 - b) 找出一条等值线起始等值点并确定判断和识别条件，追踪一条等值线的全部等值点
 - c) 连接各等值点绘制光滑曲线
- 4) 等值线绘制算法——基于不规则格网（主要是 TIN 数据）方法
 - a) 内插等值点的平面位置：对原有数据进行内插，找出等值点，等值点的内插都是在三角形的边上进行的
 - b) 找起始和终止等值点：
 - i. 具有 z 值的等值点可以组成一条以上的等值线；
 - ii. 开曲等值线：开始于制图区域的边界又结束于边界，因此，起始等值点和终止等值点一定位于边界三角形的最外边上；
 - iii. 闭合等值线：一定位于制图区域内部，其内部三角形边上任一等值点均可作为线头和线尾。
 - c) 追踪等值点并计数
 - d) 等值线注记和曲线光滑

2. 垂直剖面线法

- 1) 垂直剖面线图：表示沿水平地面一条线方向上高程的变化

3. 地貌晕渲法

- 1) 模拟太阳光与地表要素相互作用的地貌可视化方法
- 2) 基本原理：向光的山坡明亮，而背光的山坡阴暗
- 3) 主要参数：太阳方位角、太阳高度角、坡度、坡向



4. 分层设色法

- 1) 基本思想：用不同的颜色来表示不同的高程分区，表示高程渐变的过程
- 2) 对于小比例尺地图有较好的地形显示效果

5. 透视图法

- 1) 立体图可以生动逼真地描述制图对象在平面和空间上分布的形态特征和构造关系
- 2) 分析立体图，可以了解地理模型表面的平缓起伏，可以分析各个断面的状况
- 3) 可以了解研究区域的轮廓形态、变化规律以及内部结构
- 4) 主要参数：观察方位、观察角度、观察距离、z 比例系数